

Flächenhafte Schätzung mit
Classification and Regression Trees
und robuste Gütebestimmung
ökologischer Parameter in einem
kleinen Einzugsgebiet

Von Marco Patrick Schillinger

Vorgelegte Dissertation zur

Erlangung des akademischen Grades

Doktor der Naturwissenschaften der

Fakultät für Biologie, Chemie und Geowissenschaften

Universität Bayreuth

Erstgutachter: Prof. Dr. Bernd Huwe

Bayreuth, Mai 2002

Flächenhafte Schätzung mit
Classification and Regression Trees
und robuste Gütebestimmung
ökologischer Parameter in einem
kleinen Einzugsgebiet

Von Marco Patrick Schillinger

Vorgelegte Dissertation zur

Erlangung des akademischen Grades

Doktor der Naturwissenschaften der

Fakultät für Biologie, Chemie und Geowissenschaften

Universität Bayreuth

Erstgutachter: Prof. Dr. Bernd Huwe

Bayreuth, Mai 2002

Vollständiger Abdruck der von der Fakultät für Biologie, Chemie und Geowissenschaften der Universität Bayreuth zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften genehmigten Dissertation

Tag der Einreichung: 8. Mai 2002

Tag des wissenschaftlichen
Kolloquiums: 27. November 2002

Prüfungsausschuss:

Prof. Dr. C. Engels	Vorsitzender
Prof. Dr. B. Huwe	1. Gutachter
Prof. Dr. R. Herrmann	2. Gutachter
Prof. Dr. K. Bitzer	
PD Dr. K.U. Totsche	

Dank

Prof. Dr. B. Huwe für die Überlassung des Themas, für die akademische Freiheit und freundliche Unterstützung, die angenehme Arbeitsatmosphäre, das jederzeit offene Ohr und seine große Geduld

Bernd Weiblen für die Digitalisierung der Bodenschätzungsdaten

Dr. Fleck für die Überlassung der Daten der bodenkundlichen Kartierung des Weiherbachgebietes

Dr. Markus Mokry für die Überlassung der Nitratgehaltsdaten

Dem Finanzamt Bruchsal für die Bereitstellung der Bodenschätzungsdaten

Den Landwirten des Weiherbachtales für ihre Bereitschaft zur Zusammenarbeit

Allen Mitgliedern der Abteilung der Abteilung Bodenphysik für das angenehme Arbeitsklima und vielfältige Unterstützung

Dem ökologisch-botanischen Garten für seine Gastfreundschaft in den vergangenen Monaten

Herrn Michael Westerman für seine Hilfe bei der Anfertigung der Summary

Dr. Kai Totsche für die vielfältige Hilfe und Nothilfe in Computerangelegenheiten

Dr. Martin Schlather für seine intensive Betreuung

Der größte Dank gebührt jedoch meiner Frau für ihre unermessliche Geduld, ihre aufmunternden Worte, tatkräftige Hilfe und Unterstützung

**Erklärungen gemäß der Promotionsordnung für die Fakultät für
Biologie, Chemie und Geowissenschaften der Universität Bayreuth in
der Fassung der Bekanntmachung vom 1. Oktober 2001:**

Ad § 4 (6):

Hiermit erkläre ich, dass ich nicht diese oder eine gleichartige
Doktorprüfung an einer anderen Hochschule endgültig nicht
bestanden habe.

Ad § 5 (3):

Ich erkläre, dass ich die Arbeit selbständig verfasst und keine anderen
als die von mir angegebenen Quellen und Hilfsmittel benutzt habe.

Ad § 5 (4):

Ferner erkläre ich, dass ich nicht anderweitig mit oder ohne Erfolg
versucht habe, eine Dissertation einzureichen oder mich der
Doktorprüfung zu unterziehen.

Bayreuth, den 6. Mai 2002

Marco P. Schillinger

Dipl. Geoökol. Univ.

Inhaltsverzeichnis

Dank	III
Erklärungen gemäß Promotionsordnung	IV
Inhaltsverzeichnis	V
Abbildungsverzeichnis	VII
Tabellenverzeichnis	VIII
Abkürzungs- und Symbolverzeichnis	IX
Zusammenfassung	XII
Summary	XIII
1. Einführung und Zielsetzung	1
1.1 Motivation und Problemstellung	1
1.2 Stand der Forschung	3
1.2.1 Die Bodenschätzung der Finanzverwaltungen der Länder	3
1.2.2 Classification and Regression Trees	7
1.2.3 Gütekriterien	10
1.2.4 Stickstoffmodelle und räumliche Schätzung von Nitratwerten	11
1.3 Fragestellungen	12
1.4 Arbeitsziele	13
1.5 Notationen und Definitionen	13
1.6 Grundsätzliche Annahmen	15
1.6.1 Überblick	15
1.6.2 Lineares Modell	15
1.6.3 Stationarität des Zufallsfeldes	16
1.6.4 Ausreißer und kleine Fehler in der Stichprobe	16
1.6.5 Konsistenz und Validität der Stichprobe	16
1.7 Arbeitshypothesen	17
1.8 Innovative Elemente der vorliegenden Arbeit	18
2. Datengrundlage	19
2.1 Untersuchungsgebiet	19
2.1.1 Lage	19
2.1.2 Geologie	19
2.1.3 Klima	21
2.1.4 Bodentypen	21
2.1.5 Bodeneigenschaften	23
2.1.6 Vegetation und Landnutzung	23
2.2 N _{min} - Untersuchungen	24
2.3 Daten der Bodenschätzung	24
2.3.1 Bodenschätzung	24
2.3.2 Bodenschätzung im Projektgebiet	26
2.3.3 Kartierungen nach Bodenkundlicher Kartieranleitung im Projektgebiet	27
2.3.4 Korngrößenfraktionsanteile	28
2.3.5 Relationen Bodenschätzung - Bodenkundliche Kartieranleitung	28
2.3.6 Bildung einer Übersetzungsfunktion	29
2.3.7 Übersetzungsfunktion nach Weiblen	29
3. Methoden	32
3.1 Grundzüge robuster Statistik	32

3.1.1 Robuste Kenngrößen	33
3.1.2 M-Schätzer	36
3.2 Gütekriterien	37
3.2.1 Allgemeine Korrelationsschemata bei Schätzungen	38
3.2.2 Analyse der Momente von Messung und Schätzung	38
3.2.3 Analyse der Schätzabweichungen von Messwerten und Schätzwerten	41
3.2.4 Referenzverfahren	44
3.3 Kreuzvalidierung	46
3.3.1 Schätzung der Güte im Untersuchungsgebiet mit der Kreuzvalidierung	47
3.4 <i>Classification And Regression Trees</i>	47
3.4.1 Einleitung und Übersicht	47
3.4.2 Kleinste Quadrate Regression	49
3.4.3 Teilung eines Knotens	50
3.4.4 Stutzen des Maximalbaumes	51
3.4.5 Eigenschaften der <i>classification and regression trees</i>	53
3.5 Schätzung des Nitratgehaltes	55
3.5.1 Schätzung des Nitratgehaltes mit CART	55
3.5.2 Vergleichsverfahren zur Schätzung der Nitratwerte im Boden	55
3.6 Schätzung der Korngrößenfraktionsanteile	56
3.6.1 Schätzung der Korngrößenfraktionsanteile mit CART	56
4. Fallstudie 1: Schätzung von Nitratwerten im Boden	57
4.1 Ergebnisse der CART-Schätzungen	57
4.2 Vergleich der Ergebnisse	57
4.3. Zusammenfassung der Ergebnisse	73
5. Fallstudie 2: Schätzung der Korngrößenfraktionsanteile	75
5.1 Ergebnisse der Schätzungen	76
5.2 Zusammenfassung der Ergebnisse	83
6. Diskussion	85
6.1 Vergleich von CART mit einfachen Referenzverfahren	85
6.2 Gütekriterien	86
6.3 Kreuzvalidierung als Verfahren zur räumlichen Güteschätzung	86
6.4 Robuste Methoden	87
6.5 N_{\min} -Beprobungsmethode	87
7. Schlussfolgerungen	88
Literatur	91
Anhang	
Anhang 1 Arbeitsbereiche der Institutionen des Weiherbachprojektes	102
Anhang 2 Abgrenzung der Kreuzvalidierung zu Jackknife und Bootstrap	103
Anhang 3 Ergänzung Robuste Statistik	110
Anhang 4 Numerische Kodierung der Bodenschätzungsbodenarten	122
Anhang 5 Gütemaße der Schätzung der Bodenfeuchte	123
Anhang 6 Bodenartentransformationsfunktion	123
Anhang 7 Nitratstickstoffgehalte im Untersuchungsgebiet	124

Der Dissertation ist eine CD-ROM mit Daten beigelegt. Eine Übersicht der Dateien findet sich in auf der CD-ROM in der MS-Word Datei ‚Inhaltsverzeichnis CD-ROM‘

Abbildungsverzeichnis

Nr.	Abbildungsname	Seite
1.1	Schematische Darstellung der Zufallsfunktion im Untersuchungsraum	14
2.1	Lage der Gemeinde Kraichtal	19
2.2	Untersuchungsgebiet ‘Weiherbach‘	20
2.3	Bodenentwicklung auf Löss	22
2.4	Nitratmessungen im Untersuchungsgebiet	24
2.5	Ausschnitt aus dem Schätzungsbuch	25
2.6	Ausschnitt einer Schätzungskarte	25
3.1	Vergleich der Robustheitseigenschaften des Mittelwertes und des Median	35
3.2	Huber-k-Schätzer definierende ψ – Funktion mit Cut-Off-Punkt k	37
3.3	Schematische Darstellung der Wahrscheinlichkeitsdichten von Beobachtungen und Schätzung	38
3.4	Wichtige Korrelationsschemata zwischen Mess- und Schätzwerten	39
3.5	Schematische Darstellung eines Entscheidungsbaumes	48
3.6	Schematische Darstellung des Baumschnitts	54
4.1	Regressionsbaum zur Schätzung des Nitratgehaltes in der Tiefe 0-3 dm	57
4.2	Regressionsbaum zur Schätzung des Nitratgehaltes in der Tiefe 3-6 dm	58
4.3	Regressionsbaum zur Schätzung des Nitratgehaltes in der Tiefe 6-9 dm	58
4.4	Schätzwerte vs. Messwerte der Verfahren zur Schätzung von Nitratgehaltes	59
4.5	Arithmetisches Mittel der Mess- und Schätzwerte nach Tiefenstufe geordnet	60
4.6	Robuste Mittelwerte der Mess- und Schätzwerte nach Tiefenstufe geordnet	61
4.7	Standardabweichungen der Mess- und Schätzwerte nach Tiefenstufe geordnet	63
4.8	Maximale Schätzabweichung (<i>me</i>) der Schätzwerte nach Tiefenstufe geordnet	64
4.9	Wurzel der mittleren quadrierten Schätzfehler (<i>rmsr</i>) der Schätzverfahren	65
4.10	Robustes Mittel (<i>rer</i>) der Schätzabweichungen der Schätzverfahren	66
4.11	Bestimmtheitsmaß (<i>cdr</i>) der Ausgleichsgeraden	67
4.12	Achsenabschnitt der robusten Regressionsgeraden	68
4.13	Steigung der robusten Regressionsgeraden	68
4.14	Koeffizient der mittleren quadrierten Schätzabweichung (<i>comsre</i>)	69
4.15	Koeffizient der maximalen Schätzabweichung (<i>comare</i>)	70
4.16	Koeffizient des zentralen Schätzfehlers (<i>cocare</i>)	71
4.17	Koeffizient des relativen robusten Schätzfehlers (<i>corore</i>)	72
5.1	Schematisierter CART-Regressionsbaum zur Schätzung des Tongehaltes	75

5.2	Schätzwerte vs. Messwerte der Verfahren zur Schätzung des Tongehaltes	76
5.3	Schätzwerte vs. Messwerte der Verfahren zur Schätzung des Schluffgehaltes	77
5.4	Schätzwerte vs. Messwerte der Verfahren zur Schätzung des Sandgehaltes	78
5.5	Arithmetische Mittelwerte der Mess- und Schätzwerte der Korngrößenfraktionsanteile	78
5.6	Robuste Mittelwerte der Mess- und Schätzwerte der Korngrößenfraktionsanteile	79
5.7	Maximale Schätzabweichung (<i>me</i>) der Korngrößenfraktionsanteile	79
5.8	Wurzel der mittleren quadrierten Schätzabweichung (<i>rmsr</i>) der Korngrößenfraktionsanteile	80
5.9	Robustes Mittel der Schätzabweichungen (<i>rer</i>) der Korngrößenfraktionsanteile	80
5.10	Bestimmtheitsmaß (<i>cdr</i>) der Ausgleichsgeraden zwischen Mess- und Schätzwerten	81
5.11	Achsenabschnitt der robusten Ausgleichsgerade (<i>a</i>) nach Tukey (1977)	81
5.12	Steigung der robusten Ausgleichsgerade (<i>b</i>) nach Tukey (1977)	81
5.13	Koeffizient der maximalen Schätzabweichung (<i>comare</i>)	82
5.14	Koeffizient der zentralen Schätzabweichung (<i>cocare</i>)	82
5.15	Koeffizient der mittleren quadrierten Schätzabweichung (<i>comsre</i>)	83
5.16	Koeffizient der robusten Schätzabweichung (<i>corore</i>)	83
A2.1	Bootstrap Schema zur Schätzung des Standardfehlers einer Statistik	104
A2.2	Allgemeines Bootstrap-Prinzip für Ein-Stichproben-Probleme	105
A2.3	Bootstrap-Prinzip zur Schätzung des Vorhersagefehlers	106
A2.4	Methodenvergleich der Verfahren Kreuzvalidierung, Jackknife und Bootstrap	109
A3.1	Sensitivitätskurve des arithmetischen Mittels und des Medians	113
A3.2	Einflussfunktionen des arithmetischen Mittels und des Medians	114
A3.3	Huber-k-Schätzer definierende Ψ – Funktion mit Cut-Off-Punkt k	120
A3.4	Schematische Darstellung der Ψ – Funktionen verschiedener M-estimators	120

Tabellenverzeichnis

Nr.	Tabellenname	Seite
2.1	Klimadaten des Untersuchungsgebietes	21
2.2	Übersicht der Kartierungen nach BKA im Untersuchungsgebiet	28
2.3	Übersetzungsfunktion der Bodenarten nach Weiblen (1997)	31
5.1	Trennvariable und Knotenzahl der verwendeten Regressionsbäume	76
A3.1	Vergleich der Robustheitseigenschaften des Mittelwertes und des Medians	116

Abkürzungs- und Symbolverzeichnis

α	Komplexitätsparameter; mit steigenden α -Werten nimmt der Einfluss der Baumgröße $ T $ auf $R_\alpha(T)$ stetig zu
α'_k	Komplexitätsparameter für das Stutzen der Kreuzvalidierungsteilbäume (Gl. 3.51)
$\gamma^*(T, F)$	gross-error sensitivity (Gl. 3.3)
ε	Modellfehler
ε^*	allgemeiner Bruchpunkt (Gl. 3.6)
ε_N^*	Bruchpunkt einer Stichprobe
λ^*	local-shift sensitivity (Gl. 3.4)
$\eta_x(t)$	Funktion, die jedem Einflussvektor t ein Schätzwert auf der Basis der Messwerte x (Paare aus gemessenen Einfluss- und Zielgrößen) zuordnet; flächige Schätzfunktion
ρ^*	Rejection point (Gl. 3.5)
$\theta, \hat{\theta}$	Parameter einer Grundgesamtheit, (Schätzfunktion für den Parameter auf Basis der Stichprobe); Bsp. Mittelwert, Median
a	Achsenabschnitt der robusten Regressionsgeraden nach Tukey (Gl. 3.26)
b	Steigung der robusten Regressionsgeraden nach Tukey (Gl. 3.27)
BKA	Bodenkundliche Kartieranleitung
$bias$	Differenz zwischen den Mittelwerten der Messungen und der Schätzwerte (Gl. 3.13)
$^{robust}bias$	robustifizierter $bias$; (Gl. 3.14)
B-robust	Bias robust
B.-W.	Baden-Württemberg
cdr	Bestimmtheitsmaß der Regressionsgerade zwischen Schätzwerten und Messwerten (Gl. 3.25)
$cocare$	Koeffizient der zentralen Schätzabweichung (Gl. 3.30)
$comare$	Koeffizient der maximalen Schätzabweichung (Gl. 3.30)
$comsre$	Koeffizient der mittleren quadrierten Schätzabweichungen (Gl. 3.30)
$core$	Familie von Gütemaßen, die auf Koeffizienten der Schätzabweichung aufbauen (Gl. 3.28)
$corore$	Koeffizient des robusten Mittels der Schätzabweichungen (Gl. 3.30)
crm	Verhältnis der Mittelwerte der Verteilungen (Gl. 3.15)
$^{robust}crm$	robustifizierter crm (Gl. 3.16)

X

$err(\mathbf{X}, F)$	Vorhersagefehler für das gesamte Untersuchungsgebiet auf der Basis der Stichprobe \mathbf{X} für die Prognose $\eta_{\mathbf{x}}(\mathbf{u}_0)$
E_{0F}	Erwartungswert einer neuen Beobachtung $\mathbf{X}_0 = (\mathbf{U}_0, Y_0)$ mit der Wahrscheinlichkeitsverteilung F
F, \hat{F}	Wahrscheinlichkeitsverteilung der Grundgesamtheit; empirische Wahrscheinlichkeitsverteilung der Stichprobe
GLA	Geologisches Landesamt
Grp-Veg	Schätzverfahren mit Hilfe der Gruppenmittelwerte aller Messungen gleicher Kulturart; Untervarianten -CV = Kreuzvalidierungsschätzung, -RE = Resubstitutionsschätzung
Grp-Vor	Schätzverfahren mit Hilfe der Gruppenmittelwerte aller Messungen gleicher Vorfrucht; Untervarianten -CV = Kreuzvalidierungsschätzung, -RE = Resubstitutionsschätzung
Grp-Betr	Schätzverfahren mit Hilfe der Gruppenmittelwerte aller Messungen mit gleichem bewirtschaftendem Betrieb; Untervarianten -CV = Kreuzvalidierungsschätzung, -RE = Resubstitutionsschätzung
$I_C(x)$	Indikator Funktion zur Aufteilung eines Knotens bei CART (Gl. 3.40f)
IF	Influence curve, Einflussfunktion (Gl. 3.2)
k	cut-off-Punkt, meist MAD
LUFA	Landwirtschaftliche Untersuchungs- und Forschungsanstalt
MAD	Median der Medianabweichung
me	maximale Schätzabweichung (Gl. 3.20)
[m/m]	massebezogener Anteil
$nFKdB$	nutzbare Feldkapazität im durchwurzelbaren Wurzelraum
N	Stichprobenumfang
NN	Normal Null
NN	Schätzverfahren Nächster Nachbar
$NN-Veg$	Schätzverfahren Nächster Nachbar mit gleicher Kulturart
$NN-Hang$	Schätzverfahren Nächster Nachbar mit gleicher Hanglage
$P(E)$	Wahrscheinlichkeit des Ereignisses E
$Q[y, \eta]$	Fehlermaß zwischen der Beobachtung und Schätzung auf der Basis $\hat{y}_i = \eta_{\mathbf{x}}(\mathbf{u}_i)$
$R^*(\eta)$	Güte der Regressionsfunktion (Gl. 3.43)
$R_\alpha(T)$	Fehler-Komplexitätsmaß eines Baumes T (Gl. 3.49)
R_L	Resubstitutionsfehler der Stichprobe L (Gl. 3.44)
$R_{L_1, ts}$	Fehler der Teststichprobe L_1 (Gl. 3.45)
$R_{L_i, cv}$	v -facher Kreuzvalidierungsfehler (Gl. 3.46)

$R_{L_1, cv}(T_{\alpha_k})$	ν -facher Kreuzvalidierungsfehler des Baumes T_{α_k} (Gl. 3.52)
$RE^*(\eta)$	relativer Fehler (3.47)
rer	robuste Lageschätzung der Schätzabweichungen (Gl. 3.23)
$rmsr$	Wurzel der mittleren quadrierten Schätzabweichungen (Gl. 3.22)
$rmsr_{norm}$	normierte Wurzel der mittleren quadrierten Schätzabweichungen (Gl. 3.22)
SC	Sensitivity curve, Sensitivitätskurve (Gl. 3.1)
$SE(R_{L_1, cv}(T_{\alpha_k}))$	Standardfehler des Kreuzvalidierungsfehlers des Baumes T_{α_k} (Gl. 3.53)
S_N	Skalenschätzung, meist ein Vielfaches des MAD (Gl. 3.9)
t	Schätzwert für den Parameter θ , spezielle Realisation der Schätzfunktion $T_n t_l$ Wurzelknoten
t_L, t_R	linke und rechte Teilmenge
T_N	Schätzfunktion (Schätzer) für den Parameter θ auf Basis der Stichprobe $x_1 \dots x_N$ vom Umfang n
T_{max}	Maximalbaum
$ T $	Anzahl der Endknoten des Baumes T
\mathbf{x}^T	Transponierte des Vektors \mathbf{x}
U, u	Zufallsvariable U , die den Wert u angenommen hat; Einflussgröße bei Modellen; bei flächigen Schätzungen ein $1 \times p$ Vektor gemessener Werte
$X; x$	Zufallsvariable X , die den Wert x angenommen hat; Einflussgröße bei Modellen;
\mathbf{x}	Vektor gemessener Werte $\mathbf{x} = (x_1, x_2, \dots, x_N)$
$Y; y$	Zufallsvariable Y , die den Wert y angenommen hat; Zielgröße bei Modellen; bei flächigen Schätzungen ein gemessener Wert
$\hat{Y}; \hat{y}$	Zufallsvariable Y , die den Wert y angenommen hat; Zielgröße bei Modellen; bei flächigen Schätzungen der von dem Schätzverfahren geschätzte Wert
\bar{z}	Ortsvektor
\mathcal{D}, \mathcal{Z}	Untersuchungsgebiet
L	Stichprobe
$L^{(l)}$	Lernstichprobe
L_1	Teststichprobe
x_1, x_2, \dots, x_n	<u>u.i.v.</u> F : unabhängige Beobachtungen mit identischer Wahrscheinlichkeitsverteilung F
s^*	beste Teilung eines Knotens t (Gl. 3.48)
S	Menge aller möglichen Teilungen eines Knotens t
$T_1 \succ T_2 \succ \dots \succ t_1$	\succ bedeutet, dass jeder Baum T_k mit $k > n$ ein Teilbaum des Baumes T_n ist

Zusammenfassung

Simulationsmodelle ökologischer Prozesse erlangen zunehmende Bedeutung als wissenschaftliche Entscheidungsgrundlage der öffentlichen Verwaltung und der Politik.

Für die Anwendung flächenhafter Prozess- und Simulationsmodelle sind hierfür zuverlässige Eingangsinformationen bereit zu stellen. Komplexe Eingangsvariablen lassen sich teilweise nicht oder nur mit unverhältnismäßig hohem Aufwand durch flächenhafte Aufnahmen und Messungen, z. B. mit Fernerkundungsverfahren ermitteln. Ihre räumliche Bereitstellung ist nur mit Hilfe von Schätzverfahren möglich.

Manche Variablen lassen sich nicht mit modernen Interpolationsverfahren, wie z. B. Kriging-Verfahren schätzen, da die Werte keinen räumlichen Zusammenhang aufweisen. In diesem Falle können jedoch andere flächenhaft ermittelbare Variable, die mit der gesuchten Variable in einem kausalen Zusammenhang stehen, als Hilfsvariable für eine räumliche Schätzung auf Grundlage von Punktmessungen dienen.

Messungen in den experimentellen Geowissenschaften sind trotz größter Sorgfalt mit Fehlern behaftet. Hierbei handelt es sich sowohl um Rundungs- und gerätebedingte Messungenauigkeiten als auch um extreme Werte.

Robuste statistische parametrische Verfahren sind in der Lage, mit beiden Fehlertypen zurecht zu kommen. Bislang gehen fast alle vorhandenen Gütemaße von fehlerfreien Messwerten aus. Die im Rahmen dieser Arbeit entwickelten Gütemaße integrieren robuste Verfahren und führen so zu Gütemaßen, die gegenüber beiden Fehlertypen robust sind.

Classification and regression trees (CART) stellen ein Verfahren dar, mit dem sowohl nominalskalierte als auch stetige Zielgrößen auf der Basis von erklärenden Variablen geschätzt werden können. Dabei können die erklärenden Variablen unterschiedlichen Skalentypen angehören. Die *classification and regression trees* lassen sich leicht interpretieren und für die Klassifizierung neuer Objekte verwenden.

Am Beispiel des 6,3 km² großen Einzugsgebiets „Weiherbach“ im Kraichgau wird die flächenhafte Schätzung mit Hilfe von *regression trees* exemplarisch für die Variablen Nitratgehalt im Boden und Korngrößenfraktionsanteile des Bodens dargestellt. Die Ergebnisse werden mit Hilfe der weiter entwickelten Gütemaße charakterisiert und mit den Ergebnissen einfacher Referenzschätzverfahren verglichen. Es zeigt sich, dass bei beiden Zielvariablen die Schätzung mit einem *regression tree* den Referenzschätzverfahren hinsichtlich den meisten Gütemaßen mindestens ebenbürtig, wenn nicht überlegen ist. Die Schätzungen mit Hilfe der *regression trees* zeichnen sich bei beiden Zielvariablen insbesondere durch einen deutlich geringeren Bias und eine geringere maximale Schätzabweichung aus. Bei der Schätzung der Korngrößenfraktionsanteile lässt sich beim CART-Verfahren gegenüber einem von Weiblen (1997) beschriebenen Verfahren ein stärkerer linearer Zusammenhang zwischen den gemessenen und den geschätzten Werten feststellen.

Robuste Gütemaße ergänzen den Kanon bisher vorhandener Gütemaße und ermöglichen eine differenziertere Beschreibung der Qualität der Schätzungen. Insbesondere die robusten Versionen der *core*-Gütemaße sind zur Formulierung von Mindestanforderungen geeignet.

Es zeigt sich, dass weniger die Schätzverfahren, als vielmehr die qualitative und quantitative Datengrundlage die Grenzen der Leistungsfähigkeit moderner Regionalisierungsverfahren darstellen.

Summary

Simulation models of ecological processes are gaining in importance in scientific work, aiding decisions in public administration and politics.

For the application of spatial process- and simulation models we need reliable, yet high definition input information. Such inputs can be inordinately difficult to attain, and to achieve a map, estimates are required.

Whereas some inputs can be successfully estimated using modern interpolation methods such as Kriging, it becomes impossible when the inputs show less spatial correlation. In this case, auxiliary variables which show a causal correlation to the inputs can be used for spatial estimates based on point measurements.

In spite of the greatest care, measurements in experimental earth sciences are not error free, because of ,wobblers‘ and ,outliers‘, not to mention human error.

Robust statistical parametrical procedures allow one to cope with both types of error. Thus far, quality criteria assumed error free measurements. The estimation performance criteria developed in this thesis integrate robust procedures and thus allow quality measurements relatively immune to error.

Classification and regression trees (CART) represent a method of estimating nominally scaled as well as real variables, output and/or input in any combination. *Classification and regression trees* can be readily interpreted and used in classifying new objects.

In the case of the 6.3 km² catchment area ,Weiherbach‘ in Kraichgau (South West Germany) the spatial estimates are represented with the aid of *regression trees* for the two examples soil nitrogen and soil particle content according to the German soil classification system.

With the help of the newly developed quality measurements the estimates are characterised and compared to the results of simple reference estimating methods. In both examples, estimates by regression trees are at least as good as if not superior to the simple reference estimating methods regarding most estimation performance criteria, namely through marginally less bias and smaller maximum estimation error. In estimating soil particle content the CART method shows a better coefficient of determination between measured and estimated values compared to a method described by Weiblen (1997).

Robust estimation performance criteria add to the canon of established quality measurement methods and allow a more differentiated quality analysis of the estimates. In particular, the robust versions of *core*-performance criteria seem to be suitable for establishing minimal quality standards.

Finally it would seem that qualitative and quantitative inputs rather than estimation methods determine the performance of modern regionalisation methods.

1 Einführung und Zielsetzung

1.1 Motivation und Problemstellung

Vor dem Hintergrund der zunehmenden Belastung des Grundwassers mit Schadstoffen in den siebziger und achtziger Jahren des letzten Jahrhunderts wurde versucht, ökologische Prozesse mit Hilfe von Prozess- und Simulationsmodellen weitgehend deterministisch und flächendetailliert nachzubilden. Ein Ziel ist hierbei unter anderem, den Stofftransport aus Einzugsgebieten zu prognostizieren, um die Folgen von direkten menschlichen Eingriffen z. B. durch Landwirtschaft oder der Auswirkungen von anthropogen bedingten Veränderungen von Umweltbedingungen auf Ökosysteme beurteilen zu können. Für die Anwendung von Gebietsmodellen ist es erforderlich, die hierzu nötigen Eingangsinformationen der Modelle flächenhaft bereit zu stellen. Hierfür ist die Anwendung von geeigneten Regionalisierungsverfahren notwendig, die die benötigten räumlichen Informationen liefern. Dabei dienen bereits vorhandene Punkt- und/oder Flächendaten als Grundlage. Klar definierte Schnittstellen zu Geographischen Informationssystemen (GIS) und Datenbanken stellen die Verbindung zu den Prozess- und Simulationsmodellen her.

Ein neues Regionalisierungsverfahren stellen die *classification and regression trees* (CART) dar, deren Eignung als flächenhaftes Schätzverfahren am Beispiel der *regression trees* in dieser Arbeit untersucht wird.

Bei der Anwendung von Regionalisierungsverfahren ergibt sich die Notwendigkeit, verschiedene Verfahren bzw. Verfahrensvarianten hinsichtlich ihrer Leistungsfähigkeit (performance) zu bewerten. Hierzu werden in der vorliegenden Arbeit sowohl bereits bestehende, als auch neue Ansätze vorgestellt.

Ein prinzipielles Problem der geowissenschaftlichen Forschung besteht darin, dass trotz aller möglichen Sorgfalt die verfügbaren Daten fehlerbehaftet sind und eine Fehlerberichtigung – wenn überhaupt – nur sehr schwierig durchzuführen ist. Daher sind Strategien erforderlich um diesem Sachverhalt gerecht zu werden. Ein Ansatz besteht darin, Konzepte und Vorgehensweisen aus dem Bereich der robusten Statistik zu übernehmen. Daher werden an den unten dargestellten Variablen beispielhaft Elemente der robusten Statistik im Bereich der Gütekriterien auf ihre Brauchbarkeit untersucht.

Die vorliegende Arbeit ist gekennzeichnet durch zwei methodische Aufgabenfelder mit neuen Aspekten:

1. Schätzung von flächenhaften Daten aufgrund von vorhandenen Informationen mit Hilfe der *classification and regression trees*
2. Erweiterung des Kanons von Gütekriterien durch Elemente der robusten Statistik

Die vorliegenden Ergebnisse wurden im Rahmen des interdisziplinären Verbundprojektes „Prognosemodell für die Gewässerbelastung durch Stofftransport aus einem kleinen ländlichen Einzugsgebiet“ (kurz: „Weiherbachprojekt“) gewonnen, das von Instituten und Lehrstühlen der Universitäten Karlsruhe, Heidelberg, Cottbus, Bayreuth und der Staatlichen Landwirtschaftlichen Untersuchungs- und Forschungsanstalt (LUFA) Augustenberg getragen wurde. Das Verbundprojekt hatte zum Ziel, ein in sich geschlossenes Gesamtmodell als Kette von einzelnen numerisch basierten Prozessmodellen für das

hauptsächlich landwirtschaftlich genutzte Einzugsgebiet des Weiherbaches im Kraichgau aufzustellen (Plate, 1992). Die Teilprojekte und Arbeitsbereiche der einzelnen Institutionen des Weiherbachprojektes sind in Anhang 1 dargestellt. Aus der Bearbeitung dieses Projektes ergaben sich zwei Zielvariable, für die in dieser Arbeit beispielhaft geeignete Regionalisierungsverfahren ermittelt und hinsichtlich ihrer Güte mit einfachen Standardverfahren verglichen und bewertet werden. Dabei handelt es sich um die Variablen Nitratgehalt im Boden und Korngrößenfraktionsanteile des Bodens. Aus den Korngrößenfraktionsanteilen ergibt sich mit Hilfe des Bodenartendiagramms (AG Bodenkunde; 1982, 1996) die Bodenart. Diese stellt einen zentralen Bestandteil für weitere darauf aufbauende Stofftransportmodelle (Wasser, Nährstoffe, Boden, Pestizide etc.) dar. Der Nitratgehalt im Boden ist die wichtigste Größe im Zusammenhang der zunehmenden Belastung des Trinkwassers mit Nitrat.

Eine weitere wichtige Zielrichtung des Weiherbachprojektes war es, in erster Linie bereits vorhandene Messdaten zu nutzen, um den Aufwand für neue Beprobungen so gering wie möglich halten zu können. Diese Vorgehensweise hat den Hintergrund, dass der finanzielle und technisch-organisatorische Aufwand realisierbar bleibt.

Im Rahmen des Weiherbachprojektes wurden u.a. die aktuelle Anbaufrucht durch Kartierung ermittelt und die topografische Struktur durch ein digitales Höhenmodell abgebildet, so dass diese als flächendeckend bekannt vorausgesetzt werden können. Desweiteren wird im Rahmen der Bodenschätzung der Finanzverwaltungen eine räumlich hoch aufgelöste bundesweite Inventur wichtiger ertragsrelevanter Bodeneigenschaften vorgenommen und in zunehmenden Maße digital gespeichert.

Es wird davon ausgegangen, dass in naher bis mittlerer Zukunft auf Grund von Fortschritten in Wissenschaft und Technik (siehe Kap. 1.2) zumindest für das Gebiet der Bundesrepublik Deutschland Vegetation, topografische Struktur und spezielle Bodeninformationen mit hoher räumlicher Auflösung leicht zugänglich sind. Diese Flächeninformationen werden daher in der vorliegenden Arbeit als Grundlage für die flächenhafte Schätzung der Variablen Nitratgehalt im Boden und Korngrößenfraktionsanteile des Bodens heran gezogen. Kalibriert und getestet werden die Schätzverfahren mit Hilfe von punktuell vorgenommenen Messungen, die im Rahmen des Weiherbachprojektes durchgeführt wurden. Diese Vorgehensweise der Schätzung auf Grundlage von Flächendaten und Kalibrierung der Schätzverfahren auf Basis von Punktdaten stellt damit einen theoretisch praktikablen Ansatz dar, Eingangsgrößen für räumliche deterministische Stofftransportmodelle zur Verfügung zu stellen. Die vorliegende Arbeit hat zum Ziel, den praktischen Wert dieser Konzeption mit den methodisch neuen Aspekten der *classification and regression trees* und verbesserten Gütemaßen beispielhaft zu bestimmen.

1.2 Stand der Forschung

1.2.1 Die Bodenschätzung der Finanzverwaltungen der Länder¹

Das Bodenschätzungsgesetz wurde 1934 in Deutschland mit dem Ziel erlassen, eine einheitliche Besteuerungsgrundlage für Acker und Grünlandflächen in Deutschland bereit zu stellen (Rösch & Kurandt, 1991). Die Durchführung des Bodenschätzungsgesetzes als Steuergesetz liegt im Zuständigkeitsbereich der einzelnen Bundesländer, es findet keine Leitung durch das Bundesfinanzministerium statt (Engel, 1998). Der Bundesminister der Finanzen erlässt lediglich Rechtsverordnungen hinsichtlich der sogenannten „Musterstücke der Bodenschätzung“ (Engel, 1998), die als Referenzflächen dienen und insbesondere bei der Aus- und Fortbildung der Amtlichen Landwirtschaftlichen Sachverständigen eine wichtige Rolle spielen. Bislang wurde in Deutschland eine Fläche von 17 Mio. ha nach einheitlicher Methodik kartiert. In der DDR lag eine Karte der Bodenschätzung im Maßstab 1:10.000² praktisch flächendeckend für die landwirtschaftlich genutzten Flächen vor (Diemann, 1998). In Österreich werden ca. 3 Mio. ha landwirtschaftlicher Nutzfläche von der Bodenschätzung erfasst. Die Erstschätzung aller zu begutachtenden Flächen wurde dort 1973 abgeschlossen. Die österreichische Bodenschätzung beruht im wesentlichen auf dem deutschen Bodenschätzungsgesetz von 1934 (Wagner, 1998). Die Bodenschätzung stellt aufgrund ihrer einheitlichen und flächendeckenden Durchführung für Mitteleuropa eine einmalige bodenkundliche Wissensbasis dar (Kühner, 1998).

Aufgrund der Differenzierungen des Ackerschätzungsrahmens der Bodenschätzung sind die daraus resultierenden Bodenzahlen die am besten begründeten und aussagekräftigen Informationen der Bodenschätzung. Die Bodenzahlen sind hervorragend geeignet, die nutzbare Feldkapazität im durchwurzelbaren Wurzelraum (nFKdB) zu beschreiben (Harrach, 1982, 1987). In einer klimatisch einheitlichen Region besteht eine positive, nichtlineare Korrelation zwischen der nFKdB und den Bodenzahlen (Peter u. Siebert, 1998; Sauer, 1998, 1999). Weitere Untersuchungen zeigen, dass die Roherträge der Flächen und die Wertzahlen der Bodenschätzung befriedigend korrelieren (Lieberoth, 1969; Harrach, 1998).

1.2.1.1 Weiterentwicklung der Bodenschätzung und Digitalisierung

Nach Beschlüssen von drei Länderministerkonferenzen³ sollen die Punkt- und Flächendaten der Bodenschätzung digitalisiert werden (Kühner, 1998). Ziel der Digitalisierung ist die Nutzung der Bodenschätzungsergebnisse zum Aufbau eines bundesweiten Bodeninformationssystems. Der Stand der Digitalisierung ist in den einzelnen Bundesländern unterschiedlich (Wallbaum et al. 1998), da der Aufbau eines Bodeninformationssystems unter die Verwaltungshoheit der Bundesländer fällt. Bei der Digitalisierung war Niedersachsen 1998 noch führend (Kühner, 1998). Ebenso wie Sachsen (Wallbaum et al. 1998) hat jedoch auch Sachsen-Anhalt zwischenzeitlich damit begonnen, alle Daten der Bodenschätzung (Flächen- und Punktdaten) digital verfügbar zu machen. Die Digitalisierung in Sachsen-Anhalt soll bis 2003 abgeschlossen sein (Guttek et al., 1998).

¹ Eine eingehendere Darstellung der Bodenschätzung erfolgt in Kap. 2.3.1

² in Mittelgebirgslagen im Maßstab 1:5.000

³ Finanzminister April 1994, Umweltminister Mai 1993 und Innenminister Juli 1994

In Niedersachsen werden seit Mitte der 80er Jahren die Amtlichen Landwirtschaftlichen Sachverständigen in ihrem Amtsbezirk in der Profilansprache nach der modernen bodenkundlichen Nomenklatur aus- und fortgebildet (Capelle u. Ahlers, 1998). Das Formblatt für das Feldschätzungsbuch der Bodenschätzung wurde auf Anregung des Niedersächsischen Landesamtes für Bodenforschung bundesweit EDV-gerecht überarbeitet und um bodenkundliche Profilbeschreibungen ergänzt (Capelle u. Ahlers, 1998; Wallbaum et al., 1998). Es ist somit die Tendenz festzustellen, dass die Systematik der Bodenschätzung zwar beibehalten, aber durch bodenkundliche Elemente ergänzt wird. Damit wird erreicht, dass Bodenschätzungsdaten weiterhin einem einheitlichen System unterliegen und daher beim Aufbau eines bundesweiten Bodeninformationssystems eine zentrale Rolle spielen können (Pfeiffer et al., 1998).

1.2.1.2 Nutzung der Bodenschätzung in Bodeninformationssystemen

Wallbaum et al. (1998) haben in einer ausführlichen Erhebung die über steuerliche Zwecke hinausgehenden Nutzungen der Bodenschätzungsergebnisse bei Bundes- und Länderbehörden zusammengetragen und finden, dass Bodenschätzungsdaten von den für Bodenkunde und Bodenschutz zuständigen Verwaltungen bislang noch sehr spärlich für den Aufbau von Bodeninformationssystemen eingesetzt werden. Auswertungen der Bodenschätzung beschränken sich auf kleinflächige, isolierte Vorhaben. Trotz Pilotprojekten in einigen Bundesländern werden Bodenschätzungsergebnisse als Basisdaten für die Umweltverwaltungen nicht im möglichen Maße genutzt. So werden beispielsweise Bodenschätzungsdaten für die Reduzierung der Nitratauswaschung im Rahmen der landwirtschaftlichen Beratung lediglich in Hessen auf nur ca. 600 ha verwendet (Wallbaum et al., 1998).

1.2.1.3 Übersetzungsfunktionen

Seit den sechziger Jahren des letzten Jahrhunderts wird versucht, Daten der Bodenschätzung für nicht-steuerliche Zwecke zu verwenden (Arens, 1960). Ansätze zur computergestützten Übersetzung in ein zeitgemäßes bodenkundliches System existieren seit Ende der 70er Jahre (Fleischmann et al. 1979, Benne et al., 1990; Fetzer u. Schlicker, 1992; Weiblen, 1997; Kraft, 1998). Ein umfassender Übersetzungsschlüssel⁴ wurde im Rahmen des Niedersächsischen Bodeninformationssystems (NIBIS) erstellt (Fleischmann et al., 1979; Benne et al., 1990; Oelkers, 1993a). Neben einer Ableitung von Bodenarten und Bodenformen werden hierbei auch weitere bodenkundlich relevante Parameter wie Eisen-, Kohlenstoff- und Humusgehalt von den Bodenschätzungsdaten abgeleitet.

Wallbaum findet bei der Erstellung eines modular aufgebauten Übersetzungsschlüssels in Ostdeutschland gemarkungsspezifische Differenzen und Eigenheiten, die auf Unterschiede zwischen den Amtlichen Landwirtschaftlichen Sachverständigen hinweisen (Wallbaum, 1991). Einige Autoren verwenden nicht nur Daten der Bodenschätzung sondern darüber hinaus auch andere Flächendaten, so z. B. Reliefparameter (Hangposition, Neigung), und geologische Daten, denen teilweise eine erhebliche Vorhersagekraft bei der Prognose der Bodenformen zugesprochen wird (Gündra, 1992; Hoffmann, 1993; Hoffmann et al., 1993; Kleefisch u. Köthe, 1993; Ostmann, 1993). Beim Niedersächsischen Bodeninformationssystem wurde versucht, diese Variablen durch das Einbringen von Erfahrungswissen zu berücksichtigen (Kleefisch, 1993). Oelkers (1993a) stellt hierzu jedoch fest, dass sowohl die Methoden zur Erfassung des bodenkundlichen Wissens in regelbasierten Vorhersagemodellen, als

⁴ Im Rahmen der vorliegenden Arbeit wird hierfür der Begriff *Übersetzungsfunktion* verwendet.

auch die Regeln selbst häufig fehlen oder nur sehr schwer zu erfassen sind. Daher kommt Oelkers (1993b) zum Schluss, dass in dieser Hinsicht Forschungsbedarf besteht.

Bei der Erstellung einer Übersetzungsfunktion auf der Basis von Grablochbeschrieben für ein ca. 1,5 km² großes, landwirtschaftlich genutztes Lössgebiet in Hessen findet Weltner (1998), dass aus punktbezogenen Bodenschätzungsdaten (Grablochbeschriebe) qualitative Aussagen zu Bodentypen und Bodeneigenschaften abgeleitet werden können, nicht aber aus den flächenbezogenen Bodenschätzungsdaten (Klassenzeichen). Weltner (1998) stellt fest, dass zwar 90 % der verwertbaren Grablochbeschriebe in die moderne bodenkundliche Systematik überführt werden können, aber nur bei ca. 60 % der Untersuchungsfläche die Übersetzung mit eigenen Geländebefunden übereinstimmt. Aus ihrer Untersuchung kommt Weltner (1998) zum Schluss, dass bei den Übersetzungsfunktionen sowohl lokale und regionale pedogene und naturräumliche Faktoren, als auch Besonderheiten, die in der Person der Amtlichen Landwirtschaftlichen Sachverständigen und im Zeitpunkt der Schätzung begründet sind, berücksichtigt werden sollten.

Neben Hartmann et al. (1998) stellen auch Capelle und Ahlers (1998) fest, dass eine direkte Ableitung von Bodentypen und Bodenarten aus dem Klassenzeichen nicht möglich ist. Als Ursache dafür wird der Informationsverlust durch die Mittelung über das Grablochprofil bei der Festlegung des Klassenzeichens betrachtet (Capelle und Ahlers, 1998). Für die Bewertung bodenartenabhängiger Eigenschaften ist daher die Berücksichtigung der Grablochbeschriebe notwendig (Hartmann et al., 1998). Die Ableitung der Bodenarten ist nach Meinung mehrerer Autoren (Wallbaum, 1991; Hierold, 1992; Göbel et al., 1995) zuverlässiger als die Ableitung der Bodentypen von Daten der Bodenschätzung. Eine separate Eichung der Ableitungsfunktionen für jede geologische Einheit wird jedoch immer als notwendig erachtet (Zeller 1992; Wittmann, 1992; Reiche u. Schleuss, 1992).

Um dazu beizutragen, die von Oelkers (1993b) beschriebenen Wissenslücken zu schließen, wurden zwei Forschungsthemen für Diplomarbeiten vergeben: Weiblen (1997) ermittelte eine Übersetzungsfunktion zur Ableitung einer Bodenartenkarte im Untersuchungsgebiet; Kraft (1998) verwendete die Bodenschätzungsdaten, um mit Hilfe von *classification and regression trees* eine Bodenformenkarte im Untersuchungsgebiet zu erstellen. Weiblen (1997) stellt eine gute Übereinstimmung der von ihm erstellten Karten der Tiefen von 5 und 40 cm mit seinen Geländebefunden fest. Auch der im Zuge der Pedogenese zu erwartende höhere Tongehalt in 40 cm Tiefe ist auf den Karten deutlich erkennbar. Allerdings gibt es auch Hinweise auf kleine personen- und zeitbedingte Differenzen bei der Durchführung der Bodenschätzung. Die von Weiblen (1997) ermittelte Übersetzungsfunktion weist teilweise erhebliche Unterschiede zu Übersetzungsfunktionen von anderen Gebieten (Benne et al. 1990; Wallbaum, 1991) auf. Die These von Wallbaum (1991), dass Übersetzungsfunktionen für jede Landschaftseinheit getrennt ermittelt werden müssen, wurde damit bestätigt. Aufgrund der großen Streuung der Korngrößenfraktionsanteile bei den vorliegenden Wertepaaren (Bodenschätzung – bodenkundliche Kartierung) sieht Weiblen als Untergrenze die Anzahl von 9 Wertepaaren für die Übersetzung einer Bodenschätzungsbodenart an. Weitere Variablen (insbesondere Zusatzinformationen der Bodenschätzung, Geologische Karten, etc.), bei denen ein kausaler Zusammenhang angenommen wird, konnten von Weiblen (1997) in der Übersetzungsfunktion nicht berücksichtigt werden. Trotz dieser methodisch bedingten Einschränkung stellt Weiblen jedoch fest, dass die Bodenschätzung eine geeignete Grundlage für die Erstellung einer Bodenartenkarte darstellt. Seine Ergebnisse motivieren, bei der Ermittlung einer Übersetzungsfunktion Zusatzinformationen zu nutzen.

Kraft (1998) nimmt die Erkenntnisse von Weiblen auf und erstellt eine Bodenformenkarte des Untersuchungsgebietes mit Hilfe von *classification trees*. Dabei verwendet Kraft als Eingangsgrößen Bodenschätzungsdaten, geologische- und Landnutzungskarten, sowie Reliefinformationen. Kraft (1998) stellt fest, dass Bodenschätzungsinformationen eine sehr große Vorhersagekraft besitzen und dadurch bei der Ableitung der Bodenformen gute Ergebnisse, mit Ausnahme einiger Teilbereiche, erzielt werden können. Die erhaltenen Karten sind als Konzeptbodenkarten eine gute Ausgangsgrundlage für eine bodenkundliche Kartierung, da letztere dann zielgerichtet und effizient erfolgen kann. Insgesamt sind nach Ansicht von Kraft *classification trees* eine geeignete Methode zur Ableitung der Bodenform.

1.2.1.4 Kritische Punkte der Bodenschätzung

Nachfolgend genannte Punkte der Bodenschätzung werden in der Literatur als kritisch betrachtet. Aufgrund der Vorgaben der „Durchführungsbestimmungen zum Bodenschätzungsgesetz“ und der „Anweisung für die technische Durchführung der Bodenschätzung“ werden die Grenzen der Klassenflächen häufig an Grundstücksgrenzen angelehnt (Rösch u. Kurandt, 1991). Der Profilaufbau des bestimmenden Grablochs wird als repräsentativ für die gesamte Klassenfläche angesehen; der natürlichen Variabilität innerhalb einer Klassenfläche kann nur eingeschränkt Rechnung getragen werden (Hierold, 1992; Hennings, 1991). Diese Problematik besteht jedoch immer, solange eine rein binäre Zugehörigkeitsfunktion zu einer Klasse verwendet wird. Eine weitere Ursache für Inhomogenitäten innerhalb von Klassenflächen liegt darin, dass in der Schätzpraxis die Bohrungen in erster Linie aufgrund gleicher bzw. ähnlicher Ertragsfähigkeit zu Flächen aggregiert wurden und weniger wegen identischem Profilaufbau (Peter u. Siebert, 1998). Wenig untersucht ist bislang auch die zeitliche Veränderung der Bodeneigenschaften auf Grund intensiver ackerbaulicher Nutzung und der sich daraus ergebenden Abweichungen von dem in der letzten Erhebung der Bodenschätzung beschriebenen Bodenzustand. Lamp (1998) schätzt in einer Auswertung von Bodenschätzungsdaten und neueren bodenkundlichen Kartierungen in Norddeutschland den Auftrag von Kolluvienmaterial in Akkumulationslagen auf bis zu 50 cm in einem vierzigjährigen Zeitraum, so dass korrespondierend mit einer teilweise deutlichen Kappung der Profile gerechnet werden muss.

Die Nutzung der Bodenschätzung für nichtsteuerliche Zwecke wird dadurch erschwert, dass die Bodenschätzungsdaten in den einzelnen Bundesländern in sehr unterschiedlicher Weise verfügbar sind (Peter u. Siebert, 1998) und ein subjektiver Einfluss der Amtlichen Landwirtschaftlichen Sachverständigen erkennbar ist (Peter u. Siebert, 1998; Weiblen, 1997). Auch sind jüngere Schätzungen nicht immer besser als ältere Schätzungen. Insbesondere in den siebziger Jahren wurden zahlreiche Flächen mit ursprünglich unterschiedlichen Klassenzeichen, die in der Erstschätzung gegeneinander abgegrenzt waren, nachträglich zusammengefasst (Peter u. Siebert, 1998).

Der Beschreibungsschlüssel der Bodenschätzung stammt aus den dreißiger Jahren und dem damaligen wissenschaftlichen Kenntnisstand. Aus diesem Grund ist eine exakte Interpretation im Sinne der modernen bodenkundlichen Systematik schwierig (Peter u. Siebert, 1998).

1.2.2 Classification and Regression Trees

Die Grundidee der *classification and regression trees* (CART), die stufenweise Partitionierung von Daten, geht auf Sonquist et al. zurück, die 1963 das *automatic interaction detection program* (AID) vorgeschlagen hatten (Sonquist et al., 1964). Breiman et al. veröffentlichten 1984 die grundlegende CART-Monografie. Die Anwendung von *classification and regression trees* ist wesentlich vereinfacht und einem breiten wissenschaftlichen Publikum zugänglich gemacht worden, als Breiman, Olshen, Stone und Friedman zusammen mit Salford Systems Inc. ein gleichnamiges Windows-Programm des CART-Algorithmus kommerziell anboten (Wilson, 1998). Eine Auflistung der Eigenschaften der *classification and regression trees* erfolgt in Kap. 3.4.5.

Entscheidungsbäume haben in verschiedenen Disziplinen Eingang in die Wissenschaft gefunden. Anwendungsschwerpunkte liegen im Bereich der Medizin, in der Betriebs- und Finanzwirtschaft sowie in den Geowissenschaften.

Im Bereich der Medizin werden Entscheidungsbäume vorwiegend bei der Prozessaufklärung, der Variablenreduktion und der Variablenselektion verwendet. Typische Anwendungen sind insbesondere die Krankheitsdiagnose und die Prognose des Heilungserfolges (Tafeit u. Reibnegger, 1999). So fanden Mair et al. (1995) wichtige klinische Symptome, mit denen bei Patienten bei einer Klinikaufnahme ein akuter Herzinfarkt schnell diagnostiziert und entsprechende Gegenmaßnahmen eingeleitet werden können. Rechenintensive nicht-parametrische Verfahren werden bislang nur selten in der Medizin verwendet (Kuhnert et al., 2000); meist werden bei der Analyse von epidemiologischen Erhebungen logistische Regressionsverfahren (Kleinbaum et al., 1982; Freeman, 1987) verwendet, die jedoch weniger informativ und anschaulich sind als CART. Bei epidemiologischen Untersuchungen können Risikobewertung, Ausreißeridentifikation und Variablenreduktion mit Hilfe des CART-Algorithmus durchgeführt werden (Kuhnert et al., 2000). *Classification and regression trees* können hierbei auch als Voruntersuchungsmethode für eine detaillierte logistische Regression dienen. Kuhnert et al. (2000) wendeten CART in diesem Sinne bei einer Analyse der Verletzungen bei Kraftfahrzeugunfällen an.

In der Betriebswirtschaft werden *classification and regression trees* ebenfalls zur Variablenselektion und Risikoanalyse eingesetzt. Es finden sich Beispiele zur Bonitätsprüfung im Kreditgeschäft (Marais et al., 1984; Hofmann, 1990; Wagner, 1997; Teufel, 1998; Siciliano u. Mola, 2000) und zur Identifikation wichtiger Einflussvariablen auf Warenverkaufspreise (Tronstad, 1995).

Bei Anwendungen in den Geowissenschaften und in der Ökologie stehen zwei eng verknüpfte Anwendungszwecke im Vordergrund: Zum einen die Prozessidentifikation, d. h. die Herstellung eines Zusammenhanges zwischen Variablen, bei denen davon ausgegangen wird, dass sie in einem kausalen Zusammenhang stehen und zum anderen die darauf aufbauende räumliche Schätzung, wenn die erklärenden Variablen durch moderne Techniken (insbesondere durch Verfahren der Fernerkundung) flächenhaft zur Verfügung stehen (siehe nächstes Kapitel). Im Bereich der Prozessforschung untersuchten Lamon u. Stow (1999a, 1999b) mit Hilfe von CART die Einflussfaktoren von Schadstoffkonzentration in Salmoniden im Lake Michigan. Negron (1998) konnte bei einer Prozessanalyse des Käferbefalls von Douglasien mittels Klassifikationsbäumen wichtige Einflussfaktoren, die mit hohem Befall einhergingen, mit einer kreuzvalidierten Genauigkeit von 0.55-0.63 % ermitteln. Auch bei der Untersuchung von Lawinenabgängen in Skigebieten wurde *classification and regression trees* verwendet um bei einer großen Zahl (31) von potenziellen Einflussfaktoren die wichtigsten her-

auszufiltern (Davis et al., 1999). Die Vorhersage von aufwändig zu bestimmenden Zielvariablen mit Hilfe von einfachen Einflussgrößen wurde mit einem mehrstufigen CART-Verfahren durchgeführt, um die Nebelwahrscheinlichkeit zu prognostizieren. Dabei erzielten Walmsley et al. (1999) eine Genauigkeit von 0.66 und 0.96 % für das Auftreten von Nebel- bzw. Nichtnebelsituationen. Bei einem Methodenvergleich von *classification trees* mit einer logistischen Regression hinsichtlich der Mortalitätsrate von Waldbäumen wurden ähnliche Ergebnisse beider Verfahren erzielt. Dabei wurden jedoch von den beiden Verfahren unterschiedliche Kombinationen erklärender Variablen verwendet. Die Genauigkeit lag zwischen 11 und 36 % (Dobbertin u. Biging, 1998). Ein wesentlich besseres Ergebnis ermittelten Reichard u. Hamilton (1997) bei einem Vergleich von Klassifikationsbäumen mit linearen Diskriminanzanalysen in Bezug auf die Merkmalsauswahl von holzigen Pflanzen, die sich als Neophyten einbürgern konnten. *Classification and regression trees* erzielten 76 %, eine lineare Diskriminanzanalyse 86,2 % kreuzvalidierte Genauigkeit. Mit CART konnte jedoch ein anschaulicher Entscheidungsbaum zum Management gegenüber potenziellen Neophyten erstellt werden (Reichard u. Hamilton, 1997). Clark et al. (1999) gelang es mit Hilfe von CART, den Nisterfolg von Fasanen in Zusammenhang mit der verfügbaren Reviergröße zu bringen und damit den Nisterfolg als Ergebnis sich ändernder Landschaftsbedingungen zu modellieren.

Classification and regression trees stehen im Wettbewerb mit Künstlichen Neuronalen Netzen (Tafeit u. Reibnegger, 1999), die auf Kohonen (1984) zurückgehen. Nach Tafeit u. Reibnegger (1999) finden Künstliche Neuronale Netze (KNN) seit Beginn der neunziger Jahre in der medizinischen Forschung zunehmend Verwendung. Weitere Alternativen zu Entscheidungsbäumen sind traditionelle Verfahren wie Diskriminanzanalyse (Lachenbruch, 1975), Cox-Regression (Cox u. Oakes, 1984), Logistische Regression, LOGIT und PROBIT-Analyse (Fox u. Long, 1990), Expertenwissen und Expertensysteme. Tafeit u. Reibnegger (1999) stellen fest, dass die Ergebnisse der KNN im medizinischen Bereich oft besser sind als die oben genannten Verfahren. Für Bonitätsanalysen wurden in der Finanzwirtschaft von Henley u. Hand (1996) auch eine Nächste-Nachbar-Methode vorgeschlagen und u. a. mit *classification and regression trees* verglichen. Auch Heikkonen u. Varfis (1998) verwenden für die Klassifikation von urbanen Gebieten *classification trees* und halten die Verwendung von KNN und Nächste-Nachbarn-Schätzern für den genannten Zweck ebenfalls für möglich. Michel et al. (1999) stellen für die Analyse des Ausgangssignals von elektronischen Schaltkreisen ebenfalls einen Algorithmus vor, mit dem ein rekursiver Regressionsbaum mit einem Entropie-basierten Knoten-Unreinheits-Kriterium erzeugt wird.

Leistungsvergleiche zwischen den konkurrierenden Verfahren wurden von mehreren Autoren durchgeführt. So stellt etwa Howieson (1991) im ökonomischen Bereich fest, dass *classification and regression trees* zumindest vergleichbare, wenn nicht geringere Fehlerraten liefern als die LOGIT-Analyse; Marais et al. (1984) finden beim Vergleich von CART und PROBIT-Analysen im wesentlichen gleichwertige Ergebnisse. Frydman et al. (1985) erzielen bei der Klassifikation der Zahlungsfähigkeit von Unternehmen, Srinivasan u. Kim (1987) bei der Bonitätsprüfung im Kreditgeschäft mit CART geringere Fehlklassifizierungsraten als mit Linearer Diskriminanzanalyse.

Weiterentwicklungen von CART fanden in verschiedenen Bereichen statt: Splittingkriterien für die Entwicklung von Maximalbäumen (Taylor u. Jones, 1996), Stutzen der Maximalbäume mit Hilfe alternativer Kosten-Komplexitätsmaße (Tibshirani, 1996; LeBlanc u. Tibshirani, 1998), Auswahl optimaler Klassifikationsbäume (Merler und Furlanello, 1997; Efron u. Tibshirani, 1997; Denison et al., 1998; Chipman et al., 1998; Chipman et al., 2000), Tests von Entscheidungsbäumen (Morishita und

Nakaya; 1999, 2000), Einsatz von Resampling-Methoden bei instabilen Entscheidungsbäumen (Breiman, 1996a, 1998a, 1998b; Freund und Schapire, 1997; Wernecke et al., 1998), Kombination von Entscheidungsbäumen und Fuzzy-Methoden (Kiselev et al., 1997; Suarez u. Lutsko, 1999). Synthese von Entscheidungsbäumen und Künstlichen Neuronalen Netzen (Kolcz, 2000).

Breiman (1996a, 1998a, 1998b) stellt die *bagging*-Prozedur vor, bei der *resampling* Methoden mit Schätzverfahren wie z. B. Entscheidungsbäumen oder Künstlichen Neuronalen Netzen kombiniert werden, um das besondere Problem von instabilen Schätzverfahren, nämlich die meist hohe Schätzvarianz zu senken. Beim *bagging* werden Bootstrap-Wiederholungen (c. f. Anhang 2) der Original-Lernstichprobe als neue Lernstichprobe verwendet und diese anschließend aggregiert. Im Falle der Regressionsbäume werden die Ergebnisse der Entscheidungsbäume der Bootstrap-Wiederholungen gemittelt, im Klassifikationsfalle werden Mehrheitsentscheidungen getroffen. Experimente mit realen und simulierten Daten zeigen, dass *bagging* zu erheblichen Verbesserungen der Genauigkeit führt, wenn die Perturbation der Lernstichprobe zu deutlich unterschiedlichen Bäumen führt (Breiman, 1996a). Freund und Schapire (1997) schlagen den *arcing*-Algorithmus vor, bei dem die Gewichte der häufig fehlklassifizierten Fälle beim *resampling* erhöht werden und die *resampling*-Stichproben anschließend ähnlich wie beim *bagging* kombiniert werden. *Arcing* erweist sich hinsichtlich des Resubstitutionsfehlers besser als *bagging* (Breiman, 1998a).

Kolcz (2000) schlägt auf Grundlage der Ähnlichkeiten zwischen N-Tupel-Neuronalen Netzen und der Bootstrap-Resampling Version der *classification and regression trees* ein neues Verfahren vor, das im Vergleich besser abschneidet.

Eine der neuesten Weiterentwicklungen der Entscheidungsbäume ist die Kombination der Fuzzy-Set-Theorie mit den Klassifikations- und Regressionsbäumen (Kiselev et al., 1997; Suarez u. Lutsko, 1999), indem die Zugehörigkeit eines Datenpunktes zu einem Knoten mit Hilfe einer Fuzzy-Zahl beschrieben wird. Durch die Kombination kann eine deutliche Verbesserung der Schätzungen erreicht werden.

Im Bereich räumlicher Schätzungen in den Geowissenschaften gibt es mehrere Beispiele für die Anwendung von CART. McBratney et al. (2000) halten nichtlineare Methoden und die Verwendung von Hilfsvariablen für besonders wichtig. Weitere interessante Verfahren sind ihrer Ansicht nach moderne Regressionstechniken wie generalisierte lineare Modelle (GLM), generalisierte additive Modelle (GAM), CART und KNN. Die verschiedenen Techniken weisen unterschiedliche Schätzfehler auf. Hybrid-Methoden mit Geostatistik sind mächtige Verfahren zur räumlichen Schätzung, insbesondere in Einzugsgebieten und auf regionaler Ebene. Die Wahl des Schätzverfahrens ist abhängig vom Zweck der Anwendung und der erforderlichen Genauigkeit.

Plant et al. (1999) bestimmen das Ausmaß und die damit verbundenen Faktoren des Ertrages in drei Weizenfeldern. *Classification and regression trees* werden hierbei verwendet, um die Faktoren zu bestimmen, die mit der Variabilität des Ertrages einhergehen. Als Hilfsvariable werden Falsch-Farben-Luftaufnahmen der brach liegenden Felder und Bodenproben, die in einem regelmäßigen Netz mit 61 m Maschenweite entnommen wurden, verwendet. Mit Hilfe von Regressionsbäumen können ertragsrelevante Faktoren wie Bodenart, Unkrautbesatz und Phosphatgehalt bestimmt und die Felder in verschiedene Management-Zonen aufgeteilt werden. Faucher et al. (1999) kombinieren in einem zweistufigen Verfahren CART mit einem Neuro-Fuzzy-Interference System (Jang u. Sun, 1995) um im ersten Schritt relevante Vorhersage-Variablen auszugliedern und im zweiten Schritt den Wind in Küstennähe auf der Basis großräumiger atmosphärischer Variablen vorherzusagen. Dabei können mehr

als 60 % der beobachteten Varianz durch das Modell erklärt werden. McKenzie u. Ryan (1999) verwenden *regression trees*, um Bodeneigenschaften auf der Grundlage leicht verfügbarer Flächendaten zu schätzen. Als Flächendaten nutzen sie ein digitales Höhenmodell mit 25 m Auflösung und luftgestützte Fernerkundung. Das Modell wurde auf der Grundlage von 165 Bodenbeprobungen erstellt. Bei der Schätzung der Bodenprofiltiefe konnte das Modell 42 %, beim Phosphorgehalt 78 % und beim Gesamt-Kohlenstoffgehalt 54 % der Probenvarianz erklären. Sie schlagen vor, trendbereinigte Daten als Ausgangsgrundlage für Kriging-Verfahren zu verwenden. In vielbeachteten Artikeln erbringen DeFries et al. (1997, 1998) und Hansen et al. (1996, 2000) den Nachweis, dass eine weltweite Kartierung mit Hilfe von Entscheidungsbäumen möglich ist. Die Landbedeckung wird von ihnen mit Hilfe von Klassifikationsbäumen auf der Grundlage von satellitengestützten Fernerkundungsdaten abgeleitet. Dabei wird eine die gesamte Landoberfläche der Erde umfassende Karte mit einer Auflösung von 8 km erzeugt. Die Genauigkeit der Zuordnung liegt zwischen 81,4 % und 90,3 %.

Kraft (1998) findet bei der Ableitung einer Bodenformenkarte des Untersuchungsgebietes mit Hilfe von *classification trees*, dass es möglich ist, Übersetzungsfunktionen für die Bodenarten auf der Basis von Entscheidungsbäumen zu ermitteln und damit Bodenformenkarten zu erstellen. Weiter untersucht Kraft (1998) die Vorhersagekraft verschiedener Einflussvariablen. Problematisch ist, dass die Bewertung der Klassifikationsbäume und der daraus erstellten Karten nur mit bodenkundlichem Sachverstand möglich ist. Als Ursache gibt Kraft (1998) an, dass die Entscheidungsbäume schon bei geringer Variation der Variablenkombinationen oder Veränderungen des Datensatzes wenig stabil sind und die Klassifikationsregeln daher oft zufallsbedingt sind. Bei Anwendung von Expertenwissen ließen sich jedoch für die Ableitung einer Bodenformenkarte gute Ergebnisse produzieren. In Teilbereichen schlechte Übersetzungsergebnisse führt Kraft (1998) auf zu geringe Stichprobenanzahl zurück. Trotz dieser Einschränkung kommt Kraft (1998) zu dem Schluss, dass Klassifikationsbäume geeignet sind, die Bodenform auf der Grundlage komplexer Datenstrukturen flächig abzuleiten.

1.2.3 Gütekriterien

Gütekriterien stellen im Ablauf Modell – Ergebnis – Interpretation – Bewertung ein zentrales Glied dar. Trotz großer Bemühungen in der Vergangenheit im Bereich der Modellentwicklung existieren nur wenige Veröffentlichungen zur Beschreibung der Güte von Stofftransportmodellen bzw. räumlichen Schätzungen (Addiscott u. Wagenet, 1985).

Einen Überblick über quantitative Methoden, welche Unterschiede zwischen gemessenen und geschätzten Werten beschreiben, geben Loague und Green (1991) und Janssen und Heuberger (1995). Sun (1994) unterscheidet zwischen *model performance criteria* und *model rejection criteria*. Letztere ermöglichen den paarweisen Vergleich von Modellen anhand der Modellergebnisse, ersteres wird für einzelne Modelle unabhängig von anderen Modellen angewendet. Meist beziehen sich *model performance criteria* jedoch auf den Mittelwert der Messwerte, so dass sie letztlich ebenso einen paarweisen Vergleich (Mittelwert vs. Modell) darstellen. Zacharias et al. (1996) sehen drei Kategorien von Gütekriterien: Vergleich deskriptiver Statistiken, Test von Hypothesen und Analyse der Residuen.

Loague und Green (1991) stellen fest, dass neben statistischen Kriterien auch die graphische Darstellung von Modellergebnissen wichtig ist. Addiscott und Whitmore (1987) konstatieren, dass ein einziges Kriterium irreführend sein kann, mehrere Kriterien zusammen jedoch das Maß der Übereinstimmung von Schätz- und Messwerten befriedigend beschreiben können. Viele Autoren (Loague u. Green, 1991; Comfort et al., 1993; Janssen u. Heuberger, 1995) bestätigen diese Meinung, indem sie

darauf hinweisen, dass jedes Kriterium spezifische Güteinformationen liefert. Janssen u. Heuberger (1995) empfehlen, Gütekriterien bei Datensätzen anzuwenden, die entweder unabhängig vom Kalibrierdatensatz sind oder sich erheblich von diesem unterscheiden, z. B. durch Anwendung der Kreuzvalidierung (engl. *cross-validation*, siehe Kap. 3.3).

Parish und Smith (1990) betonen, dass der Kern einer Modellvalidierung darin bestehe, ein quantitatives Gütekriterium in Hinsicht auf die Modellanwendung zu definieren und die Modellergebnisse damit zu beurteilen.

Es besteht allgemein Übereinstimmung darin, dass eine Beurteilung von Modellen aus graphischen Darstellungen und quantitativen Methoden bestehen soll (Donigian und Rao, 1990; Pennell et al., 1990; Loague und Green, 1991). Dabei kann eine graphische Darstellung helfen, Anomalien und Unterschiede zwischen den geschätzten, und den gemessenen Werten aufzudecken und Unterschiede zwischen mehreren Modellergebnissen herauszuarbeiten (Zacharias et al., 1996).

Alewell u. Manderscheid (1998) heben hervor, dass bei kleiner Varianz der Messwerte auch Modelle mit einer negativen *model efficiency* die Messwerte gut beschreiben können. Sie schlagen den *confidence interval test* (CIT) vor, mit dem jedoch lediglich der Anteil von Schätzwerten außerhalb eines Vertrauensbereiches der Messwerte beschrieben werden kann.

Die Kriterien von Loague und Green (1991) können als Vergleich des mittleren Schätzfehlers des zu untersuchenden Modells mit dem mittleren Schätzfehler eines Referenzmodells interpretiert werden. Als Referenzmodell wird hierbei ein geeigneter konstanter Wert verwendet. Hierfür kommt zunächst der Mittelwert der Messwerte in Frage. Dieser ist jedoch nicht robust (im Sinne von Huber [1981]) gegenüber Ausreißern, so dass bereits ein einziger falscher Messwert das Gütekriterium beliebig beeinflussen kann (siehe Kap. 3.1), und dadurch die Aussagekraft des Gütemaßes sehr stark eingeschränkt wird. Auf Robustheitsaspekte von Gütekriterien gehen Zacharias et al. (1996) ein. Sie schlagen vor, nichtrobuste Lokationsmaße der Gütekriterien von Loague u. Green (1991) durch den Median zu ersetzen. Dieser Ansatz hat zwar den Vorteil, robust gegenüber Ausreißern zu sein, ist aber hinsichtlich des Einflusses von kleinen Fehlern (sog. ‚Wackler‘; siehe Kap. 3.1) sogar erheblich sensibler als der üblicherweise verwendete Mittelwert. Daher besteht also trotz der Weiterentwicklung von Zacharias et al. (1996) eine methodische Lücke im Bereich eines Referenzschätzers, der sowohl gegenüber großen Ausreißern, als auch kleinen Messwertfehlern robust ist und damit die Eigenschaften von Huber (1981) erfüllt.

1.2.4 Stickstoffmodelle und räumliche Schätzung von Nitrat-Werten

Die ersten Versuche, mit Modellen komplexe ökosystemare Zusammenhänge zu beschreiben, reichen bis in die zwanziger Jahre des letzten Jahrhunderts zurück und basieren auf gekoppelten Differenzialgleichungen (Lotka-Volterra-Modelle; Lotka, 1925; Volterra, 1926). Zur Stickstoffdynamik in Böden gibt es unterschiedliche und immer detaillierter werdende Modelle und Simulationsansätze. Stickstoffs simulationsmodelle sind gekoppelt an Modelle des Wasserhaushaltes im Boden. Einfache Plattenmodelle zur Beschreibung des Bodenwasserhaushaltes liefern unter bestimmten Umständen (z.B. auf leichten, nicht zu trockenen Böden) ganz brauchbare, mit Richards-Modellen vergleichbare Ergebnisse (Wegehenkel, 1996; Richter et al., 1996; Götz-Huwe et al., 1989). Auf schwereren Böden dagegen ist der Einsatz von Modellen, die auf der Richardsgleichung basieren, erforderlich, um genauere Ergebnisse zu erzielen. Dazu ist jedoch je nach den herrschenden Randbedingungen der Rechenaufwand und

die Datenerhebung sehr aufwändig. In Thres et al. (1998) und Mertens (2000) sind Zusammenfassungen über Modelltypen und Modelltypen verschiedener Dimensionalität dargestellt. Stofftransportmodelle für ungesättigte Böden gibt es erst in neuerer Zeit, z.B. WHNSIM (Huwe, 1992), Minerva (Kersebaum et al., 1987) oder Hydrus 1D (Vogel et al., 1996) und Hydrus 2D (Simunek, Šejna, van Genuchten, 1999). Ein ausführliches Review bzw. eine Übersicht über Stickstofftransportmodelle ist bei Engel et al. (1993) bzw. Stark et al. (1997) zu finden. Einige neuere Modelle sind für die Simulation von Flusseinzugsgebieten entwickelt worden: DAISY/MIKE-SHE (Styczen & Storm, 1993, Refsgaard et al., 1999), CWSS (Reiche, 1994), CATCHN (Cooper et al., 1994), NMS (Lunn et al., 1996), AgriFlux (Banton & Larocque, 1997), INCA (Whitehead et al., 1998), SHETRAN (Birkinshaw & Ewen, 2000a, 2000b). Die Entwicklungstendenz der Stickstoffmodelle geht hin zu einer sehr detaillierten Betrachtung der verschiedenen Stickstoff- und Kohlenstoffpools und deren komplexen Transformationsprozessen (Banton & Larocque, 1997; Birkinshaw & Ewen, 2000b). Da der Stickstoffgehalt eine sehr hohe räumliche Variabilität im Boden aufweist (Huwe, 1992, Schmidhalter et al. 1992, Illsemann et al., 2001) wird als Anfangsbedingung für räumliche Simulationen vielfach mit gemittelten Daten gerechnet, zu echten Regionalisierungsverfahren der Stickstoffgehalte des Bodens finden sich nur wenige in der Literatur beschriebene Ansätze.

Reiche (1991) kalibriert sein Modell mit Hilfe von Messungen und Simulationen an Einzelstandorten, verwendet aber für die flächenhafte Simulation eines Teileinzugsgebietes von 2,3 km² Größe einheitliche Werte ohne deren Festlegung näher zu begründen. Birkinshaw und Ewen (2000b) führen repräsentative Messungen der in ihrem Einzugsgebiet vorkommenden drei Vegetationstypen Acker, Weide und Wald durch, um Anfangsbedingungen für ihre Simulation mit SHETRAN zu erhalten. Ebenso beproben Tychon et al. (1999) lediglich Standorte mit den wichtigsten Anbaufrüchten in ihrem Einzugsgebiet von 0,32 km². Huwe und Totsche (1995) hingegen klassifizieren schon homogene Einheiten hinsichtlich Bodeneigenschaften und Feldfrucht in Verbindung mit Messungen charakteristischer Kombinationen dieser beiden Parameter. Lasserre et al. (1999) schließlich versuchen, die räumliche Verteilung der Eingangsgrößen des Modells AgriFlux mit Monte Carlo Simulationen stochastisch abzubilden.

1.3 Fragestellungen

Im Rahmen dieser Arbeit sollen insbesondere folgende Fragestellungen bearbeitet werden:

- Inwiefern können *regression trees* als räumliche Schätzverfahren eingesetzt werden?
- Wie können bestehende Gütemaße ergänzt werden um Gütemaße zu erhalten, die robuste Eigenschaften im Sinne von Huber (1981) aufweisen?
- Wie ist die Performance der *classification and regression trees* im Vergleich zu einfachen Schätzverfahren?
- Wie lassen sich Nitratgehalte im Boden mit Hilfe der *classification and regression trees* im Boden schätzen?
- Auf welche Weise lässt sich eine Übersetzungsfunktion für die Schätzung der Bodenarten auf Grundlage der Bodenschätzung der Finanzverwaltung mit Hilfe der *classification and regression trees* ermitteln?
- Mit welcher Genauigkeit werden die Variablen bei der zur Verfügung stehenden Datengrundlage räumlich geschätzt?

1.4 Arbeitsziele

Das Ziel der Arbeit besteht darin,

1. den bestehenden Kanon von Gütemaßen um robuste Elemente zu erweitern
2. diese an zwei Fallbeispielen zu studieren
3. die Variablen Nitratgehalt im Boden und Bodenart im Untersuchungsgebiet flächenhaft zu schätzen. Die Schätzung soll auf der Grundlage von flächig und punktuell vorliegenden Daten mit *classification and regression trees* durchgeführt werden
4. die Ergebnisse der Schätzungen mit den *classification and regression trees* mit den Ergebnissen von sehr einfachen Referenzschätzverfahren zu vergleichen. Hierbei sollen oben genannte Gütemaße eingesetzt werden

1.5 Notationen und Definitionen

An dieser Stelle folgt eine kurze Einführung in die Grundlegenden Notationen und Definitionen, die im folgenden Kapitel verwendet werden. Kleine fett gedruckte Buchstaben - \mathbf{x} - kennzeichnen Vektoren, z. B. $\mathbf{x} = (x_1, x_2, \dots, x_N)$, fett gedruckte Großbuchstaben - \mathbf{X} - stehen für Matrizen; nicht fette Großbuchstaben beziehen sich auf eine Zufallsvariable. Fette, kursive Großbuchstaben bedeuten, dass ein Vektor von Zufallsvariablen vorliegt: $\mathbf{U} = (U_1, U_2, \dots, U_p)$. Die Transponierte eines Vektors oder einer Matrix wird mit einem hochgestellten T gekennzeichnet: \mathbf{X}^T . Parameter der Grundgesamtheit werden durch griechische Buchstaben bezeichnet: θ . $P(E)$ steht für die Wahrscheinlichkeit des Ereignisses E . Ein Dach auf einem Buchstaben, wie $\hat{\theta}$, signalisiert eine Schätzung, ein Pfeil hingegen einen Ortsvektor: \vec{z} . I_C ist die Indikator-Funktion, die den Wert 1 annimmt, wenn die Bedingung C wahr ist und 0 wenn die Bedingung falsch ist. Der Ausdruck x_1, x_2, \dots, x_n u.i.v. F bedeutet, dass die Beobachtungen unabhängig voneinander sind und eine identische Wahrscheinlichkeitsverteilung F aufweisen. Eine Schätzfunktion $T_N = T_N(x_1, \dots, x_N)$ basiert auf einer Stichprobe x_1, \dots, x_N vom Umfang N . Der Grenzwert einer Schätzfunktion T_N für $N \rightarrow \infty$ wird mit T gekennzeichnet.

Gegeben sei eine Zufallsfunktion $\{X(\vec{z}_i), \vec{z}_i \in \mathcal{G}\}$ in einem Untersuchungsgebiet \mathcal{G} . Die der Zufallsfunktion zu Grunde liegenden Zufallsvariablen $X_i = X(\vec{z}_i)$ seien jeweils unabhängig und identisch verteilt (u.i.v.) mit einer unbekannten Wahrscheinlichkeitsverteilung F aus dem Zahlenraum $H = \mathbb{R}^{(p+1)+}$ (Efron, 1983):

$$X_1, X_2, \dots, X_n \text{ u.i.v. } F \text{ mit} \quad (1.1)$$

$$F(\vec{z}; x) = P\{F(\vec{z}) \leq x\} \quad (1.2)$$

Es liegen N Realisationen, d. h. beobachtete bzw. gemessene Werte im Untersuchungsgebiet vor:

$$X_1 = x_1, X_2 = x_2, \dots, X_n = x_N \quad (1.3)$$

Die Zufallsvariable X_i am Messpunkt $\vec{z}_i \in \mathcal{G}$ sei dabei ein Vektor $\mathbf{X}_i = (\mathbf{U}_i, Y_i)^T$. \mathbf{U}_i stellt dabei den $1 \times p$ Vektor der Prädiktor-Variablen dar, welche durch Kartierung, Fernerkundung etc. flächenhaft bekannt sei: $\mathbf{U}_i = (U_{i1}, \dots, U_{ip})$. Y_i ist die flächig gesuchte, realwertige Zielvariable, deren Realisatio-

nen y_i an den Messpunkten $\bar{z}_i \in \mathcal{S}$ durch punktuelle Messung bekannt sei. Die unbekannten Realisationen y_0 sollen für jeden Punkt \bar{z}_0 im Untersuchungsgebiet mit Hilfe eines Schätzverfahrens

$$\hat{y}_0(\bar{z}_0) = \eta_{\mathbf{x}}(\mathbf{u}_0) = \eta(\mathbf{u}_0, \hat{F}) \quad (\text{nach Efron, 1993, verändert}) \quad (1.4)$$

nur auf Basis der Stichprobe \mathbf{X} mit der Stichprobenverteilung \hat{F} und dem Prädiktorvektor \mathbf{u}_0 am Schätzort \bar{z}_0 geschätzt werden. Das Schätzverfahren ist daher nicht von der Stichprobenzahl N abhängig. (Efron, 1983). Es liegt also ein univariates Problem vor, da die Zielvariable Y_i eindimensional ist. Abb. 1.1 verdeutlicht den Zusammenhang zwischen Zufallsvariable, -funktion und Einfluss- und Zielgröße.

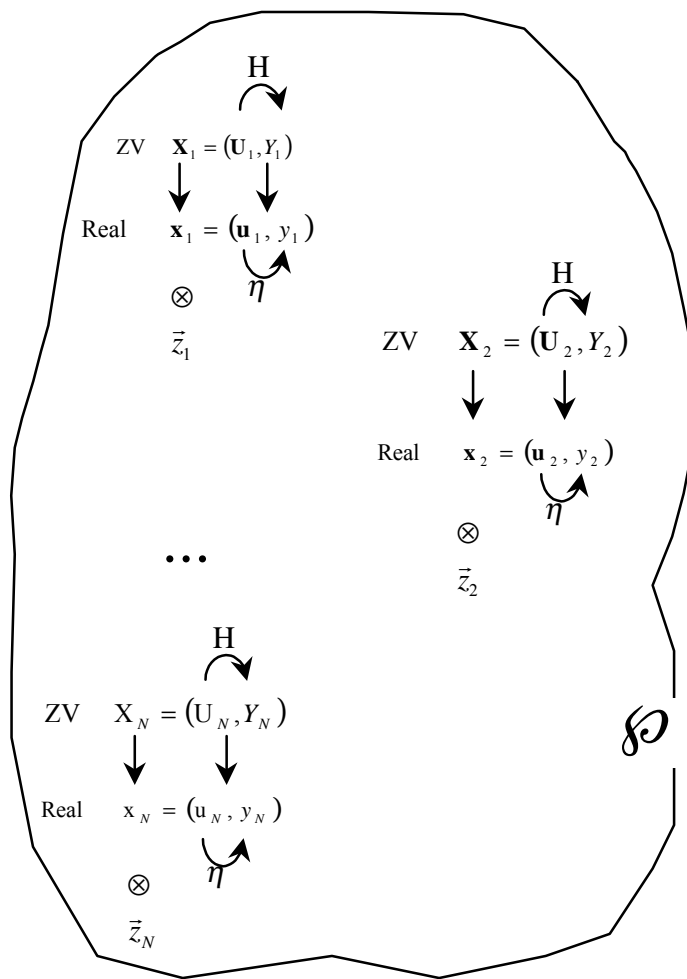


Abb. 1.1: Schematische Darstellung der Zufallsfunktion im Untersuchungsraum

Die Messungen bzw. Beobachtungen \mathbf{x}_i sind Realisationen (Real) der Zufallsvariable (ZV) \mathbf{X}_i an den Stellen \bar{z}_i (durch \otimes gekennzeichnet) im Untersuchungsgebiet \mathcal{S} . Das Set aller Zufallsvariablen \mathbf{X}_i in \mathcal{S} stellt die Zufallsfunktion $\{\mathbf{X}(\bar{z}_i)\}$ dar. Es besteht ein Zusammenhang H zwischen P flächig bekannten Prädiktorvariablen \mathbf{U}_p und der zu schätzenden Variable Y . Die Zufallsvariablen \mathbf{X}_i setzen sich daher aus dem Prädiktorvariablenvektor $\mathbf{U}_{ip}(\bar{z}_i)$, $p = 1 \dots P$, und der erklärten Variable $Y(\bar{z}_i)$ zusammen. Ziel ist es, aus den Realisationen \mathbf{x}_i ein optimales Modell η (Schätzverfahren) zwischen den Prädiktorvariablen und der erklärten Variablen zu erstellen, um letztere flächig schätzen zu können.

Eine wichtige Größe der Untersuchungen stellt das Fehlermaß $Q[y, \eta]$ zwischen der Beobachtung y_i und der Prognose $\hat{y}_i = \eta_{\mathbf{x}}(\mathbf{u}_i)$ dar. Im Falle einer Regression wird oft das euklidische Distanzmaß

$$Q[y, \eta] = (y - \eta)^2 \quad (\text{Efron, 1993}) \quad (1.5)$$

verwendet, im Klassifikationsfall einer nominalskalierten Variablen ist

$$Q[y, \eta] = I_{\{y \neq \eta\}} \text{ d.h. } \begin{cases} 1 & \text{falls } y \neq \eta \\ 0 & \text{falls } y = \eta \end{cases} \quad (\text{Efron, 1993}) \quad (1.6)$$

ein sinnvolles Maß für die Übereinstimmung zwischen Messung und Schätzung. Der Vorhersagefehler für das gesamte Untersuchungsgebiet auf der Basis der Stichprobe \mathbf{X} für die Prognose $\eta_{\mathbf{x}}(\mathbf{u}_0)$ ist definiert durch

$$err(\mathbf{X}, F) := E_{0F} \left\{ Q \left[Y_0, \eta_{\mathbf{x}}(U_0) \right] \right\} \quad (\text{Efron, 1993}) \quad (1.7)$$

Dabei bezeichnet E_{0F} den Erwartungswert einer neuen Beobachtung $\mathbf{X}_0 = (U_0, Y_0)$ mit der Wahrscheinlichkeitsverteilung F im Untersuchungsgebiet \wp (Efron, 1993). In Abb. 1.1 ist die Zufallsfunktion schematisch dargestellt.

1.6 Grundsätzliche Annahmen

1.6.1 Überblick

Die nachfolgend dargestellten Verfahren beruhen auf verschiedenen Eigenschaften der zu untersuchenden Grundgesamtheit und der hieraus durch Messung oder Beobachtung entnommenen Stichproben. Da die geforderten Eigenschaften nicht strikt bewiesen werden können, sondern nur begründete Hinweise auf ihre Existenz vorliegen, werden fünf grundsätzliche Annahmen getroffen:

1. Eingangs- und Zielgröße können mit Hilfe eines linearen Modells beschrieben werden
2. Stationarität des Zufallsfeldes
3. Die Stichproben können Ausreißer enthalten
4. Die Stichprobenwerte können kleinen Fehlern unterliegen
5. Validität und Konsistenz der Stichprobe

Nachfolgend werden diese Annahmen erläutert.

1.6.2 Lineares Modell

Es liege ein Zusammenhang zwischen den Eingangsgrößen \mathbf{U} und der Zielgröße Y vor, der mit einem linearen Modell wie folgt beschrieben werden kann: Es existiere ein $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ so, dass für alle $\mathbf{x}_i = (\mathbf{u}_i, y_i)$ im Untersuchungsgebiet \wp , $\wp \in \mathfrak{R}^{p+1}$ gelte (Efron, 1993):

$$y_i = \mathbf{u}_i \cdot \boldsymbol{\beta} + \varepsilon_i \quad (1.8)$$

wobei die Modellfehler ε_i eine unabhängige und zufällige Stichprobe einer Zufallsvariable mit einer unbekannten Fehlerverteilung F mit dem Mittelwert Null sind (Efron, 1993):

$$E_F(\varepsilon) = 0 \quad (1.9)$$

1.6.3 Stationarität des Zufallsfeldes

Das lineare Modell treffe im gesamten Untersuchungsgebiet auf gleiche Weise zu, d. h. das lineare Modell sei invariant gegenüber Translationen \vec{t} der Beprobungsorte \vec{z}_i der Beobachtungsmatrix \mathbf{X} :

$$\hat{\beta}(x(\vec{z}_1), x(\vec{z}_2), \dots, x(\vec{z}_n)) = \hat{\beta}(x(\vec{z}_1 + \vec{t}), x(\vec{z}_2 + \vec{t}), \dots, x(\vec{z}_n + \vec{t})) \quad (1.10)$$

für alle \vec{t} : $\vec{z}_i + \vec{t} = \vec{z} \in \mathcal{D}$ für alle \vec{z}_i .

1.6.4 Ausreißer und kleine Fehler in der Stichprobe

In der wissenschaftlichen Praxis treten oft Abweichungen von strengen parametrischen Modellvorstellungen über die Verteilungen von Messwerten auf (Hampel et al., 1986). Die Abweichungen können verschiedene Ursachen haben: Das Auftreten von Fehlern bei der Datengewinnung und -aufbereitung (Messung, Aufschrieb, Runden, Gruppieren, Fehlerfortpflanzung, Berechnung etc.). Derartige Fehler können sowohl ein größeres (sog. Ausreißer) als auch kleineres Ausmaß annehmen und in der Menge guter Daten verborgen bleiben (Hampel, 1980). Ein einziger nicht bemerkter Fehler kann jedoch eine statistische Analyse vollständig verderben, insbesondere wenn ein kleinsten-Quadrat-Algorithmus verwendet wird (Hampel, 1980). Ein Anteil von mehreren Prozent grober Fehler ist in der Wissenschaft verbreitet (z. B. Cochran, 1947; Daniel, 1976; Daniel & Wood, 1971).

1.6.5 Konsistenz und Validität der Stichprobe

Den Begriff der *Suffizienz einer Beprobung* wird in dieser Schrift in Anlehnung der Definition der *Suffizienz einer Stichprobenfunktion/Statistik* von Sachs (1997 S. 127 VII), Manoukian (1986 S. 22) und Rinne (1995) verwendet: Eine Stichprobenfunktion $\hat{\theta}(X_1, \dots, X_n)$ heißt suffizient, wenn die gesamte Information über den gesuchten Parameter θ , die in den beprobten Zufallsvariablen (X_i) steckt, sich auch in der Stichprobenstatistik $\hat{\theta}$ wiederfindet (Rinne, 1995 S. 88 III)

Es wird angenommen, dass das Funktional T einer Schätzfunktion T_N *Fisher-konsistent* ist (Kallianpur & Rao, 1955), d. h. dass die Folge der Funktionale $\{T_N; N \geq 1\}$ asymptotisch die richtige Größe schätzt (Hampel et al., 1986):

$$T_N \xrightarrow{N \rightarrow \infty} T(F) \quad (1.11)$$

Den Begriff *Validität der Stichprobe* wird nach Sachs (1997) verwendet. Unter Validität einer Stichprobe ist der Grad definiert, mit dem eine Beobachtung dasjenige Merkmal der Untersuchungseinheit [d. h. im Untersuchungsgebiet] charakterisiert, das der Beobachter zu erfassen wünscht (Lienert, 1973), im konkreten Fall heißt dies, dass die Beobachtung alle relevanten Einflussgrößen umfasst und die gesamte Information über das gesuchte Modell $y_i = \mathbf{u}_i \cdot \boldsymbol{\beta} + \varepsilon_i$, die in dem Untersuchungsgebiet steckt, sich auch in den Beobachtungen $\mathbf{x}_i = (\mathbf{u}_i, y_i)$ wiederfindet.

Der Begriff der Validität einer Stichprobe geht insofern über die Eigenschaft der Repräsentativität einer Stichprobe hinaus, als die Stichprobe/Messung zusätzlich zur Repräsentativität alle wesentlichen Informationen und Sachverhalte der Grundgesamtheit über die gesuchte Zielgröße bzw. den Zusammenhang der Einfluss- und der Zielgrößen erfassen soll. Daraus resultiert die Forderung nach einer quantitativ und qualitativ ausreichenden Stichprobennahme.

Nach Lienert (1973) ist die *Reliabilität* eine notwendige, aber nicht hinreichende Voraussetzung der Validität. Darunter versteht man die Reproduzierbarkeit (Sachs, 1997) hinsichtlich der ausgewerteten Beobachtungsreihen unter gleichgebliebenen Bedingungen, wobei Objektivität eine notwendige Voraussetzung der Reliabilität darstellt (Lienert, 1973).

Die Verlässlichkeit der aus Stichproben gezogenen Schlussfolgerungen hängt von der Repräsentativität der Stichprobe ab (Lienert, 1973). Repräsentativität kann durch eine zufallsgesteuerte Beprobung erreicht werden, indem jedes Element der Population die gleiche Chance hat, in die Stichprobe aufgenommen zu werden. Haben die Untersuchungselemente nicht die gleiche Chance, in die Stichprobe aufgenommen zu werden, entstehen u. U. verzerrte (biased) Stichproben, die nicht repräsentativ für die Grundgesamtheit sind und die daher keine oder nur bedingte Schlüsse auf letztere zulassen. In der empirischen Forschung kommen häufig anstelle von Zufallsstichproben sogenannte Ad-hoc-Stichproben vor, nämlich wenn Daten aus bestimmten Gründen gerade zugänglich sind. Wenn überhaupt, so kann man von solchen Stichproben lediglich auf eine fiktive Population schließen, aus der die Ad-hoc-Stichprobe dann eine Zufallstichprobe darstellt (Lienert, 1973). Bei repräsentativen Stichproben stellt die Häufigkeitsverteilung \hat{F} der Stichprobe eine gute Annäherung an die Verteilung der Grundgesamtheit F dar, d. h. die Stichprobe kann die jeweilige Grundgesamtheit bis auf zufällige Messfehler vertreten (Sachs, 1997) und ist in allen Verteilungskennwerten ein übereinstimmendes Abbild der Grundgesamtheit (Lienert, 1973).

1.7 Arbeitshypothesen

Aus dem Stand der Forschung ergeben sich folgende Arbeitshypothesen:

- Der Nitratstickstoffgehalt im Boden ist räumlich inhomogen und nur in geringem Ausmaß autokorreliert. Die flächige Schätzung des Nitratgehaltes mit bekannten Methoden ergibt keine wesentlich bessere Schätzung als bei Verwendung des Gebietsmittelwertes.
- Zur Beurteilung von Regionalisierungsverfahren sind objektive Maßzahlen erforderlich.
- Bisher angewandte Kriterien zur qualitativen Beurteilung von flächigen Schätzungen sind wenig geeignet um Aussagen über die Verlässlichkeit der Verfahren zu treffen.
- Die Kreuzvalidierung ist eine brauchbare Methode um Abschätzungen der Güte von Schätzverfahren im Raum zu erhalten.
- Robuste Statistik ist ein hilfreiches Instrument um mit fehlerbehafteten Daten umgehen zu können. Mit neu zu entwickelnden robusten Gütekriterien lässt sich der Kanon der klassischen Gütekriterien qualitativ erheblich erweitern.
- Robuste Gütekriterien sind besser geeignet um bei fehlerbehafteten Daten die Güte von Schätzungen zu beurteilen als nichtrobuste.
- Robuste Kriterien bewirken in Kombination mit nichtrobusten Kriterien einen Erkenntnisgewinn bei der Gütebeurteilung.
- Robuste Gütekriterien führen zu schärferen Aussagen hinsichtlich der Güte von Schätzverfahren.
- Die *classification and regression trees* können für räumliche Schätzungen verwendet werden.
- Mit Hilfe der *classification and regression trees* können bestimmte ökologische Parameter besser geschätzt werden als mit herkömmlichen Methoden.
- *Classification and regression trees* sind robuster als einfache Vergleichsverfahren.

1.8 Innovative Elemente der vorliegenden Arbeit

Folgende Punkte der Arbeit sind neu:

- Entwicklung von robusten Gütemaßen zur Beurteilung von Modellergebnissen unter der Annahme, dass Messwerte fehlerbehaftet sind
- Anwendung der *classification and regression trees* zur räumlichen Schätzung der Korngrößenfraktionsanteile des Bodens und des mineralischen Stickstoffgehaltes im Boden
- Anwendung voll robuster Gütemaße bei der Schätzung
- Der Ansatz, robuste Verfahren auch im Bereich von Gütekriterien anzuwenden und so zu einer differenzierteren, verlässlicheren Grundlage zur Gütebeurteilung zu kommen

2 Datengrundlage

2.1 Untersuchungsgebiet

2.1.1 Lage

Das Untersuchungsgebiet befindet sich im Kraichgau-Hügelland zwischen den Städten Karlsruhe, Heilbronn und Heidelberg im nordwestlichen Teil Baden-Württembergs im Zentrum des Kraichgaus (Abb. 2.1); es befindet sich auf dem Verwaltungsgebiet der Stadt Kraichtal auf den Gemarkungen der drei Stadtteile Menzingen, Münzesheim und Gochsheim.

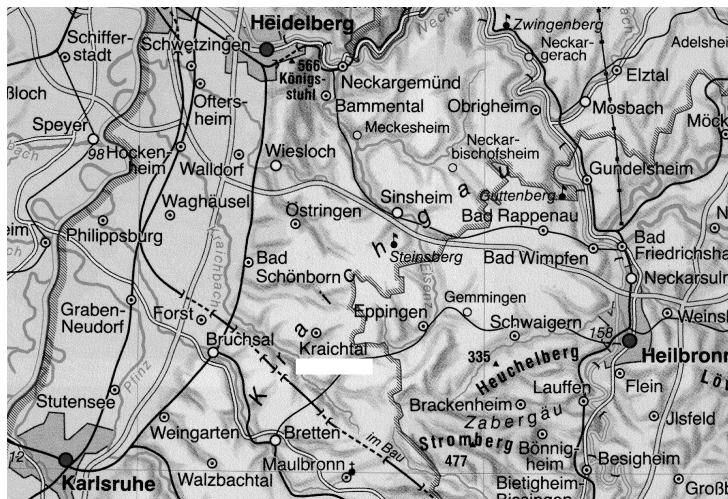


Abbildung 2.1: Lage der Gemeinde Kraichtal

Die Lage der Gemeinde Kraichtal ist hell unterstrichen. Ausschnitt aus Diercke (1988)

Das Untersuchungsgebiet (Abb. 2.2) entspricht dem oberirdischen Einzugsgebiet des Weiherbachs (Hötzl u. Ackermann, 1996, 1999), welcher am südlichen Ende des Untersuchungsgebietes in den Kraichbach mündet, der zum Rhein hin entwässert. Die Größe des Weiherbacheinzugsgebietes beträgt 6,3 km², die Höhenlage reicht von 142 bis 243 m über NN. Das bearbeitete Gebiet stellt ein typisches Beispiel eines landwirtschaftlich genutzten Kleineinzugsgebietes in einer Lösslandschaft in Mitteleuropa dar.

2.1.2 Geologie

Der Kraichgau stellt aufgrund seiner geologischen, geomorphologischen, petrographischen und klimatischen Gegebenheiten eine eigenständige Landschaftseinheit in Südwestdeutschland dar. Er wird durch seine Lage in einer tektonischen und morphologischen Senke im Norden durch den Odenwald, im Süden durch den Schwarzwald, im Westen durch den Oberrheingraben und im Osten durch das Neckartal begrenzt (Schaar, 1989). Die Kraichgausenke ist dem süddeutschen Schichtstufenland zuzurechnen, das von flach nach Ost bis Südost einfallenden mesozoischen Sedimentgesteinen geprägt wird (Henningsen u. Katzung, 1997). Der geologische Untergrund des Kraichgaus wurde in der Trias, hauptsächlich während der Epoche des Keupers, abgelagert. Die Anlage des Reliefs erfolgte im Tertiär (Eitel, 1989), seine heutige Prägung erfuhr das Gebiet während des Pleistozäns durch Lössseinwehungen.

Im Einzugsgebiet des Weiherbaches bilden Gipskeuper, Schilfsandstein und Rote Mergel des mittleren Keupers den Untergrund. Die Sedimente des Gipskeupers und der Roten Mergel stellen sich als Wechsel von tonigen rot-violetten bis grau-grünen Mergeln mit unterschiedlichen Kalkgehalten dar

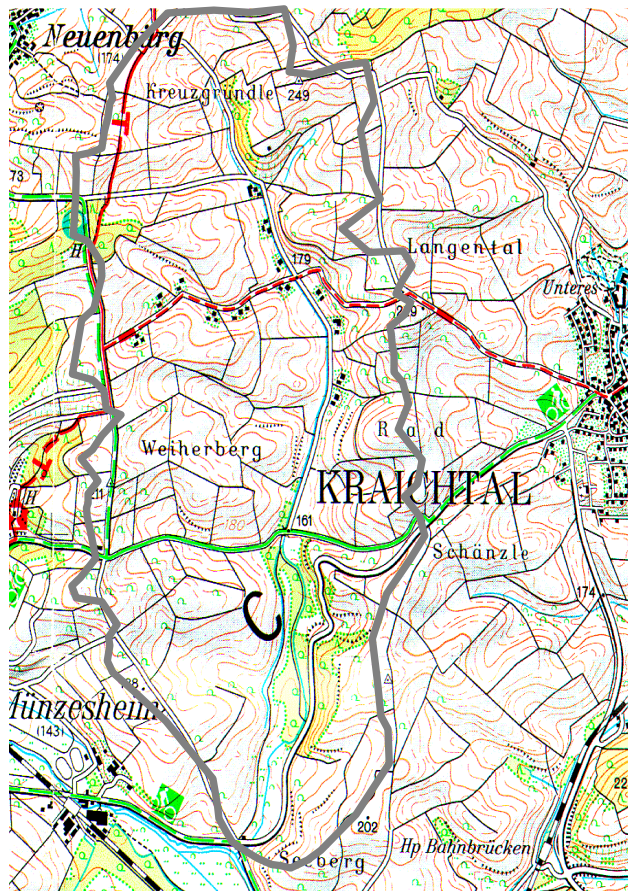


Abbildung 2.2: Untersuchungsgebiet 'Weiherbach'

Das Untersuchungsgebiet ist grau umrandet. Ausschnitt aus der Topographischen Karte 1:25.000 Blatt 6818 Kraichtal (Landesvermessungsamt Baden-Württemberg, 1996)

(Schweizer & Kraatz, 1982). Aufgrund der mächtigen Lössüberlagerungen auf der Westseite des Weiherbachtals stehen die Keuperschichten nur an den Steilhängen auf der Ostseite des Tales an. Schilfsandstein tritt nur in zwei kleinen Vorkommen im nördlichen Weiherbachtal auf. Er besteht aus rotbraunem, grobsandigem Material, das im Untersuchungsgebiet mit roten bis grauen, tonigen Mergellagen durchsetzt ist (Geologisches Landesamt Baden-Württemberg, 1985).

Schon im höheren Jura setzte, durch eine Aufwölbung im Bereich des späteren Rheingrabens verursacht, eine Kippbewegung der mesozoischen Sedimente der süd-deutschen Großscholle nach Osten ein (Pflug, 1982). Diese Kippbewegung der Sedimente wird in Verbindung mit rückschreitender Erosion als Ursache für die Anlage der asymmetrischen Talformen des Kraichgaus angesehen (Eitel, 1989). Die Osthänge im Untersuchungsgebiet sind mit nur 3° mittlerem Gefälle auffallend flacher als die Westhänge mit 8° - 11° mittlerem Gefälle. Die Asymmetrie der Talformen begünstigte im Pleistozän die Ablagerung

von Lösssedimenten v. a. im Leebereich der Osthänge. Die Sedimente stammen hauptsächlich aus der westlich gelegenen Schotterebene des Rheingrabens, wo es während der Glaziale zu starken Winderosionsprozessen kam. Die Sedimentation erfolgte anschließend vorzugsweise im Bereich der morphologischen Muldenlage des Kraichgaus (Geyer u. Gwinner, 1991). Spätere kleinräumigere Erosionsprozesse im Untersuchungsgebiet führten zu einer Umverlagerung der Lössdecke, wobei an den steileren Westhängen das anstehende Gestein teilweise wieder freigelegt wurde, während sich das erodierte Material in Senken, am Hangfuß und im Talgrund ablagerte. Als Folge dieser Umverteilung und der ursprünglichen ungleichmäßigen Sedimentation schwankt die Lössmächtigkeit im Untersuchungsgebiet zwischen Null und mehr als 20 m (Mauder, 1992). Im Weiherbachgebiet finden sich nur Lösssedimente aus dem Würm- und Rißglazial (Hötzl et al. 1992). In die Lösspakete sind geringmächtige und nicht durchgehend erhaltene Lösslehmschichten in Form von reliktschen und fossilen Bt-Horizonten eingeschaltet, die aus interstadialen Bodenbildungen stammen. Die beiden Würm- und Rißsedimente selbst sind durch eine bis zu vier Meter mächtige tonreiche Lösslehmschicht voneinan-

der getrennt. Diese Lösslehmschicht stellt den Rest einer fossilen Bodenbildung aus dem Riß-Würm-Interglazial dar (Mauder, 1992).

Bedingt durch die Erosionsanfälligkeit der Lössüberdeckung finden im Holozän zum Teil beträchtliche Hangabtragungen statt. Die holozänen Sedimente sind im gesamten Weiherbachtal als Kolluvien in Form von kalkhaltigem Schwemmlöss und entkalktem Schwemmlösslehm mit Mächtigkeiten bis 2 m anzutreffen. Auensedimente finden sich lediglich im südlichen Weiherbachtal (Mauder, 1992).

2.1.3 Klima

Der Kraichgau weist eine mittlere Jahrestemperatur von etwa 10° C, mit mittleren Januartemperaturen um 0° C und mittlere Julitemperaturen von 17° - 18° C auf (Eitel, 1989). Die mittleren Jahresniederschläge betragen zwischen 750 mm und 850 mm (Deutscher Wetterdienst, 1953). In den Sommermonaten kommt es regelmäßig zu Gewittern mit Niederschlägen über 75 mm, die starke Erosionsereignisse auslösen können (Beudert, 1997). Die Vegetationsperiode beträgt 230-240 Tage. Der Kraichgau ist mit oben genannten Klimadaten als klimatischer Gunstraum anzusprechen (Schaar, 1989). Tab. 2.1 fasst einige wichtige Klimadaten des Untersuchungsgebietes zusammen.

Tabelle 2.1: Klimadaten des Untersuchungsgebietes

Mittelwerte der Jahre 1991-1995; Die maximale Niederschlagsintensität wurde mit 10 min Intervallen bestimmt, die klimatische Wasserbilanz ist auf das hydrologische Jahr bezogen (1.11.-31.10), Daten von Kolle u. Fiedler (1999)

Lufttemperatur in 2 m Höhe	10.3° C
Bodentemperatur in 1 cm Tiefe	9.8° C
Niederschlag	785 mm
Maximale Niederschlagsintensität	23-103 mm/h
Aktuelle Evapotranspiration	564 mm
Klimatische Wasserbilanz	157 mm

naten kommt es regelmäßig zu Gewittern mit Niederschlägen über 75 mm, die starke Erosionsereignisse auslösen können (Beudert, 1997). Die Vegetationsperiode beträgt 230-240 Tage. Der Kraichgau ist mit oben genannten Klimadaten als klimatischer Gunstraum anzusprechen (Schaar, 1989). Tab. 2.1 fasst einige wichtige Klimadaten des Untersuchungsgebietes zusammen.

2.1.4 Bodentypen

Im Bereich des Westhanges des Weiherbaches haben sich meist Löss-Erosionscatenen (siehe Abb. 2.3) herausgebildet (Wagner, 1991), bei denen Pararendzinen, erodierte Parabraunerden und Kolluvisole aus Löss vorherrschen. Die Ursache für die Herausbildung von degradierten Böden liegt in dem verstärkten Bodenabtrag aufgrund der landwirtschaftlichen Nutzung.

Die bei ungestörter Bodenbildung auf Löss zu erwartenden Parabraunerden (Eitel, 1989) finden sich heute allerdings selbst in relativ ungestörten Bereichen, wie unter Wald, nur noch selten (Fleck, 1994; Gerold et al., 1992). Im Kraichgau kommen als erste Stufe der Erosionsentwicklungsreihe unter Acker höchstens erodierte Parabraunerden mit einem 4-5 cm mächtigen Bt-Horizont unter dem Ap-Horizont vor (Fleck, 1994). Bei fortgesetztem Abtrag entwickeln sich insbesondere an den Oberhängen der Lössrücken Pararendzinen (Dikau, 1986). Diese sind flach bis mittel tief und über das ganze Profil karbonatreich. Die Ap- und IC-Horizonte weisen eine stark schluffige Textur auf. Das erodierte Material akkumuliert im Hangfußbereich und in den Auen (Schulte, 1995). Die Mächtigkeit der Kolluvisole nimmt generell vom Mittelhang- über den Unterhang bis in den Senkenbereich fortlaufend zu. Aufgrund ackerbaulicher Maßnahmen und Eingriffe der Flurbereinigung finden sich jedoch auch in Oberhang-, Kuppen- und Plateaulage Kolluvisole (Kraft, 1998). Bork u. Ricken (1983) weisen auf die Ab-

hängigkeit verschiedener Bodenkennwerte von der Hangposition hin. Dabei ergeben sich die Bodenkennwerte in regelhafter Weise aufgrund der erosionsbestimmten Bodenabfolgen. Dies betrifft in erster Linie Textur und Humusgehalt, aber auch Nährstoffausstattung, Bodenwasserhaushalt und biologische Aktivität.

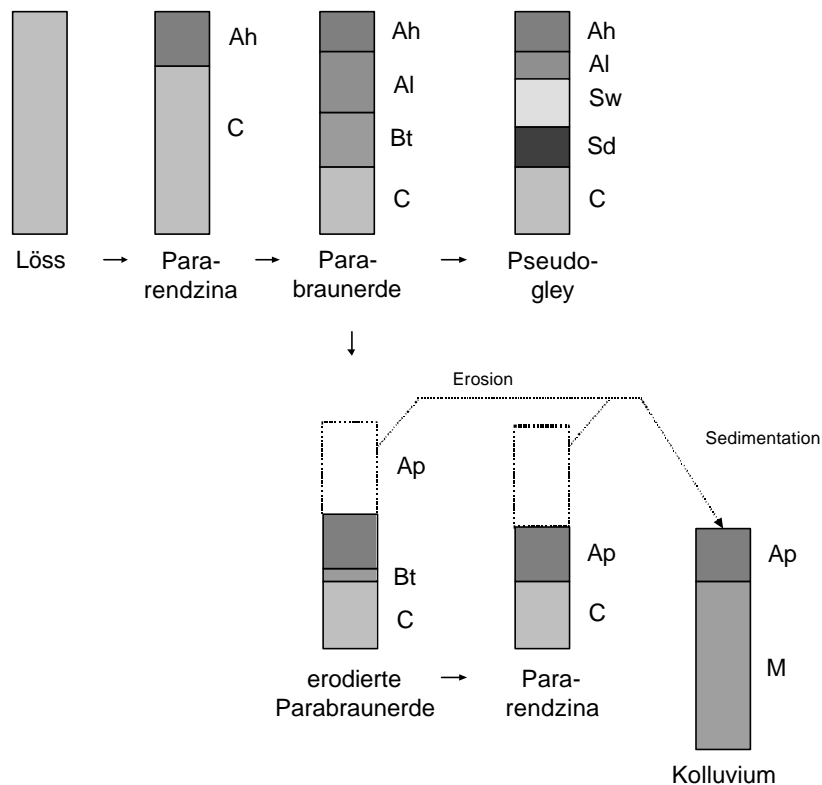


Abbildung 2.3: Bodenentwicklung auf Löss

Im oberen Block ist die Entwicklung auf ungestörten Standorten dargestellt, im unteren Block die Degradation unter Erosionseinfluss; verändert nach Schäfer (1999)

In den Lössschichten kommen dunkel- bis rötlichbraune, tonreiche (bis über 30 % [m/m]) Lösslehm-schichten vor, die als Paläoböden bezeichnet werden. Dabei handelt es sich um die Bt-Horizonte von interglazialen und interstadialen Parabraunerden. Die Parabraunerden wurden zu Beginn der auf das Interglazial oder Interstadial folgenden Kaltzeit zuerst bis auf die Bt-Horizonte erodiert und während des darauffolgenden Hochglazials mit neuen Lösssedimenten überdeckt. Im Untersuchungsgebiet wurden vier verschiedene alte Paläoböden gefunden, die aber nicht alle flächendeckend vorhanden sind (Hötzl u. Ackermann, 1998). Die Paläoböden wurden aufgrund der nutzungsbedingten Erosion teilweise wieder freigelegt. Aufgrund unterschiedlicher Erosionswirkung und ungleichem Erhaltungsgrad der Lösslehme ergibt sich ein fast zufälliges Muster von Lösslehminseln innerhalb der dominierenden Lössüberdeckung. Die Paläoböden treten hauptsächlich in den Mittelhangbereichen auf, da die Hänge aufgrund der Kolluvien im Unterhangbereich rezent eine höhere Erosionsbasis und somit eine geringere Hangneigung aufweisen als zu den Zeiten der ursprünglichen Bodenbildungen. Eine detaillierte Kartierung eines Lösshanges im Untersuchungsgebiet ist bei Delbrück (1997) zu finden.

Im Bereich der steileren Osthänge des Weiherbachtals sind die Verhältnisse aufgrund der Vielfältigkeit der Substrate und der Nutzungsgeschichte als Weinberge komplizierter. So finden sich vornehmlich Pararendzinen in Kuppenbereichen und an Steilhängen, sowie Braunerde-Pelosole unter Wald auf erosionsgeschützten Keuperstandorten. Rigosole kommen vor allem an Keupersteilhängen mit Südwest-Exposition unter rezenten oder früheren Weinanbauflächen vor. Im Bereich der Unterhänge, Senken und terrassierten Bereiche der Keupersteilhänge sind Keuper-Kolluvisole anzutreffen (Wag-

ner, 1991). Insgesamt bedecken Böden aus Keupermaterial aber weniger als 10 % der Fläche des Weiherbachgebietes (Schäfer, 1999). In der Talaue im Süden des Untersuchungsgebietes kommen Auen-
glye vergesellschaftet mit allochthonen Kalkvegen vor (Weiblen, 1997).

2.1.5 Bodeneigenschaften

Die Lössböden im Untersuchungsgebiet weisen eine Korngrößenverteilung mit einem engen Maximum meist im Grobschluffbereich (Schulte, 1995; Eitel, 1989) auf. Gerold et al. (1992) publizieren einen Grobschluffanteil von 50-55 % und eine Porosität von 0.45-0.5. Der Karbonatgehalt im Weiherbachgebiet schwankt zwischen 25-30 % [m/m]. Durch den hohen Karbonatgehalt werden die Mineralpartikel bei Trockenheit stabilisiert (Pécsi u. Richter, 1996), dadurch erklärt sich die Stabilität der früher weit verbreiteten Hohlwege. Im feuchten Zustand neigen die Böden jedoch zur Verschlammung und sind bei Oberflächenabfluss leicht erodierbar (Gerlinger, 1996). Starkregenereignisse, die zu Oberflächenabfluss führen, sind daher trotz ihres seltenen Auftretens für den weit überwiegenden Teil des Bodenabtrags verantwortlich (Beudert, 1997). Eine detaillierte Darstellung der Erosionsdynamik ist bei Gerlinger (1996) zu finden.

2.1.6 Vegetation und Landnutzung

Das Untersuchungsgebiet, dessen natürliche Vegetation aus einem lichten Eichenmischwald bestand (Firbas, 1949) wurde schon seit dem Neolithikum besiedelt (Schottmüller, 1961; Bienwald, 1970) und aufgrund seiner klimatischen, pedologischen und geomorphologischen Eigenschaften bevorzugt ackerbaulich genutzt (Walter, 1986); der ursprüngliche Waldbestand war schon vor 5.000 Jahre größtenteils gerodet (Wagner, 1991), die heutige Verteilung von bewaldeten und waldfreien Gebieten hat sich wahrscheinlich schon bis zum Ende des 15. Jahrhunderts herausgebildet (Schaar, 1989).

In den Jahren von 1962-1969 fand auf den Gemarkungen Menzingen und Münzesheim eine Flurbereinigung statt. Hierbei wurden Hohlwege verfüllt, Terrassenböschungen eingeebnet, der vormals mäandrierende Weiherbach begradigt und das Bachbettniveau teilweise abgesenkt (Mauder, 1992).

Die heutige Landnutzung weist mit einem Flächenanteil von nur 5 % vereinzelte Waldvorkommen an den Keupersteilhängen und im Auenbereich des Weiherbaches auf. Obstplantagen, Weinberge und Dauergrünland kommen im Untersuchungsgebiet nur sporadisch vor, Ackerbau hat eine dominante Stellung mit einem Flächenanteil von 90 % (Bronstert, 1994). Dabei werden hauptsächlich Getreide (Winterweizen, Wintergerste) und Hackfrüchte (Mais, Zuckerrüben) angebaut (Timmermann et al., 1998). Es herrschen mineralische Düngung und Gülledüngung (Schweinegülle) vor. Festmist hat einen vernachlässigbaren Anteil (Timmermann et al., 1998).

2.2 N_{min}-Untersuchungen

Im Rahmen des Weiherbachprojektes wurde von der LUFA Augustenberg von 1990 - 1995 ein Monitoring des mineralischen Stickstoffgehaltes im Untersuchungsgebiet durchgeführt. Dabei wurden an bis zu 51 Standorten (Abb. 2.4) innerhalb des Arbeitsgebietes 3-4 mal jährlich die Bodenschichten 0-3 dm, 3-6 dm und 6-9 dm beprobt und der Nitratgehalt im Boden nach LUFA-Richtlinien (Hoffmann et al., 1991) ermittelt. Es wurden Mischproben aus drei Wiederholungen in den Eckpunkten eines

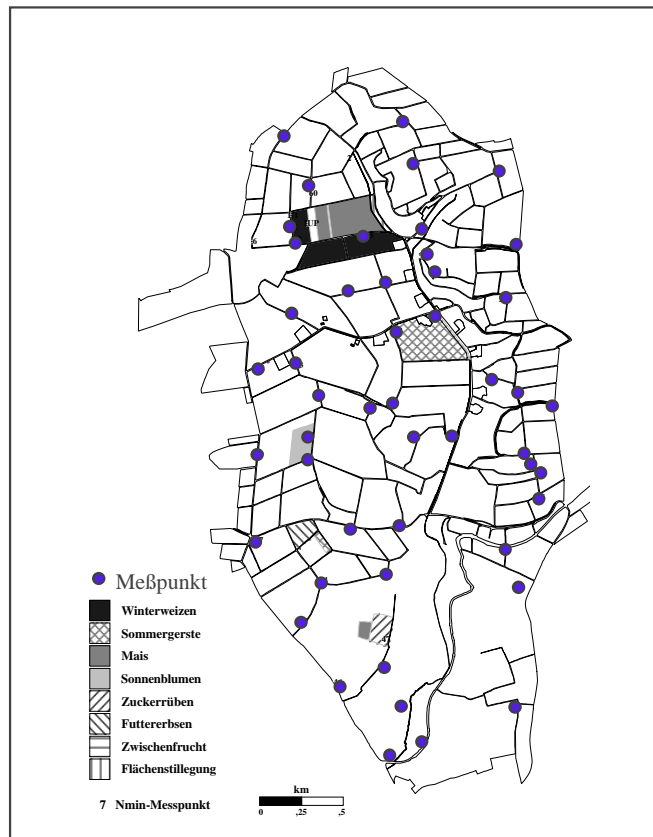


Abb. 2.4: Nitratmessungen im Untersuchungsgebiet
Quelle: LUFA Augustenberg

gleichseitigen Dreiecks mit 10-15 m Grundseite gebildet. Als zusätzliche Landnutzungsdaten wurde die aktuelle Kulturart, die Vorfrucht, das Düngungsregime und der bewirtschaftende Betrieb des beprobten Schläges und die Hangposition der Probenahmestelle erhoben. Des weiteren wurde in den genannten Jahren jährlich die Hauptkulturart flächendeckend kartiert. Die oben genannten Daten wurden von der LUFA Augustenberg zu Verfügung gestellt. Die in dieser Schrift vorgestellten methodischen Ansätze zur flächenhaften Schätzung der Nitratwerte im Boden wurden beispielhaft auf den Datensatz der Beprobung vom Februar 1992 angewendet, deren Daten in Anhang 7 dokumentiert sind. Bei dem genannten Beprobungstermin wurden auf 48 Schlägen Proben entnommen und untersucht.

2.3 Daten der Bodenschätzung

2.3.1 Bodenschätzung

Die Ziele der Bodenschätzung sind eine gerechte Bodenbesteuerung, sowie eine tragfähige Planungsgrundlage der Bodennutzung und -beleihung (Freund, 1992; Rau, 1992). Hierzu wird eine genaue Kennzeichnung des Bodens nach seiner Beschaffenheit angestrebt, um mit diesen Informationen die potenzielle Ertragsfähigkeit objektiv bestimmen zu können. Neben der Aufnahme der Bodenansprache in einem *Schätzungsbuch* (Abb. 2.5) erfolgt die Kartierung einheitlicher Bodenflächen in einer *Schät-*

zungskarte (Arens, 1960; Abb. 2.6). Die Durchführung der Bodenschätzung wird durch die "Durchführungsbestimmung zum Bodenschätzungsgesetz" (BodSchätzDB) sowie durch die „Anweisung für die technische Durchführung der Bodenschätzung“ (BodSchätzTechnAnw) näher geregelt (Rösch &

Schätzungsbuch für Ackerland												Seite 8/8
Gemeindebezirk <i>Goldschleim</i>			Fruchtigkeitszustand des Bodens <i>gut</i>			Tag der Schätzung <i>11.9.50</i>			Zugeschnitt <i>XX-V</i>			
Bemerkung <i>Flur</i>			ha, davon Grünland <i>21</i> ha									
Grablöcher			Feststellungen über Klassenflächen, Klassenabschnitte und Sonderflächen									
Nr.	Bodengefüge	Boden- zahl	Boden- art	Boden- stufe	Boden- festung	Boden- zahl	Besonderheiten	± n.S.	W- gemeines Klima ± n.S.	Bewer- zung	Erklär- ungen	Bemerkungen
1	2	3	4	5	6	7	8 bis 12	13	14	15	16	17
13	L 4 L 4 4 4 4 L 4 L 4 5 2 grü r 1/2 1/2	57	4	4	✓	58	✓ gel ver	-4 -6	6-50			
14	L 4 L 4 2 L 4 L 3 p 1/2 1/2	68	14	3		68	✓ gel	± 0	10	25		

Abbildung 2.5: Ausschnitt aus dem Schätzungsbuch

Dargestellt sind zwei ‚bestimmende‘ Grablochbeschreibungen im Untersuchungsgebiet. Bestimmende Grablöcher sind für eine Klassenfläche typisch, nach ihnen wird die Klassenfläche klassifiziert.

landschätzrahmens vergeben. Beide Schätzrahmen besitzen den gleichen Aufbau, jedoch eine andere Gewichtung der relevanten Parameter. Im folgenden wird lediglich der Ackerschätzungsrahmen kurz skizziert, da im Untersuchungsgebiet nur vernachlässigbare Grünlandanteile vorliegen. Die Böden werden durch den Ackerschätzungsrahmen an Hand der drei Merkmale *Zustandsstufe*, *Bodenart* und *Entstehung* bestimmten Klassen zugeordnet, die wiederum eine gewisse Wertspanne aufweisen, um

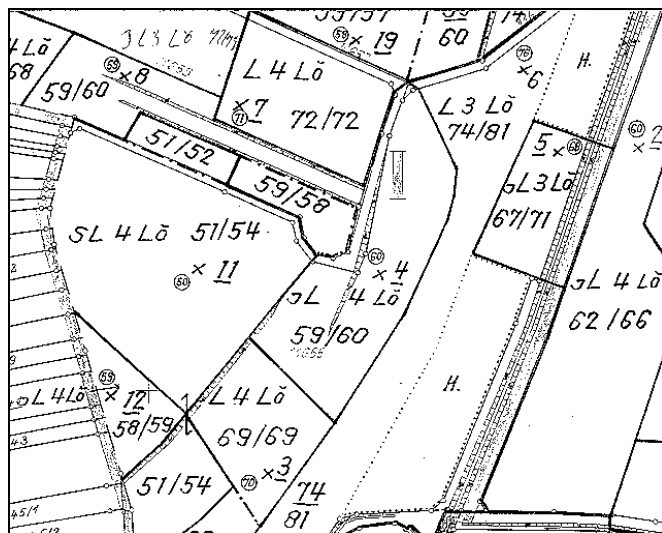


Abb. 2.6: Ausschnitt einer Schätzungskarte

Anteilen von Sand (> 10 µm) und Ton (< 10 µm) unterschieden. In die Karte aufgenommen wird dabei die durchschnittliche Bodenart des ersten Meters des Bodenprofils. Das Merkmal *Entstehung* trägt einer geologischen Komponente Rechnung, die nur auf dem Alter, und der Lagerung (fest oder locker)

Kurandt, 1991). Die Bodenschätzung wird nur bei den Nutzungsarten Acker-, Garten- und Grünland angewendet. Die Böden werden je nach potenzieller Ertragsfähigkeit durch Wertzahlen zwischen 7 und 100 charakterisiert, wobei dem Boden mit der besten Ertragsfähigkeit in Deutschland die Bewertungszahl 100 zugeordnet wird. Die Bewertungszahlen werden anhand eines Acker- und eines Grün-

auch innerhalb einer Klasse je nach Ausprägung differenzieren zu können. Die *Zustandsstufe* beschreibt diejenigen Qualitätsmerkmale der Böden, die durch die Bodenbildung und Bodenentwicklung einer zeitlichen Änderung unterliegen. Die Stufe eins kennzeichnet dabei den günstigsten, die Stufe sieben den ungünstigsten Zustand. Die Einteilung der *Bodenarten* erfolgt nach Thaer in acht mineralische Bodenartengruppen und eine Moorgruppe (Rösch & Kurandt, 1991). Die mineralischen Gruppen werden im Gegensatz zur heute üblichen Einteilung nach Kartieranleitung nur nach

des Bodenmaterials sowie der Art seiner Ablagerung und Sortierung durch Wind, Wasser oder Eis beruht (Arens, 1960). Es werden Diluvium (D; ältere Bezeichnung für Pleistozän), Alluvium (Al; ältere Bezeichnung für Holozän), Löss (Lö) und Verwitterungsböden (V) unterschieden. Die mit dem Schätzungsrahmen ermittelte *Bodenzahl* bringt nur Ertragsunterschiede zum Ausdruck, die auf die Verschiedenheit der *Bodenbeschaffenheit* zurückzuführen sind. *Bodenunabhängige* Faktoren wie Klima, Geländegestaltung oder Beschattung durch Wald, die für die Ertragsleistung mitentscheidend sind, werden durch den Schätzungsrahmen nicht berücksichtigt. Sie finden erst bei der Ermittlung der *Ackerzahl* Berücksichtigung, indem entsprechende Zu- und Abschlüsse auf die Bodenzahl gebildet werden.

Die zusammenfassende Darstellung der oben genannten Bodenkennwerte bezeichnet man als *Klassenzeichen*, das sowohl im Schätzungsbuch, als auch in der Schätzungskarte vermerkt ist. Als Beispiel bedeutet das Klassenzeichen „L3Lö 76/84“ die Bodenart Lehm mit der Entwicklungsstufe 3 aus Löss, mit einer Bodenzahl von 76 und einer Ackerzahl von 84.

Die Kartierung im Gelände erfolgt im Raster mit max. 50 m Maschenweite durch Bohrstockbeprobungen. Flächen mit einheitlichem Klassenzeichen werden als *Klassenflächen* bezeichnet, die durch ein für die Fläche typisches *Grabloch* (sog. ‚bestimmendes Grabloch‘) näher charakterisiert werden (BodSchätzTechnAnw Ziff 6). Die Kartierungsergebnisse werden sowohl in der Schätzungskarte als auch im Schätzungsbuch festgehalten, wobei die Angaben im Schätzungsbuch wesentlich detaillierter sind (z. B. Profilaufbau, Bodenartcharakterisierung, Standortcharakterisierung) als die Angaben der Schätzungskarte. In Abb. 2.5 sind zwei *Grablochbeschriebe* aus dem Schätzungsbuch, in Abb. 2.6 ist ein Ausschnitt der Schätzungskarte dargestellt.

Die Bodenschätzung wurde in Deutschland auf allen Acker- und Grünlandstandorten durchgeführt und wird auch heute bei größeren Eingriffen in die Kulturlandschaft, z. B. bei einer Flurbereinigung, aktualisiert und somit auf dem laufenden Stand gehalten.

In der DDR erfolgte 1954 eine Reorganisation der Bodenschätzung mit deren Überleitung vom Innenministerium zum Ministerium für Land- und Forstwirtschaft (Fischer, 1992). Die Bodenschätzung war zu diesem Zeitpunkt flächendeckend durchgeführt, wurde aber danach nicht mehr aktualisiert, so dass alle heute vorliegenden Daten auf dem Stand von 1955 sind. Wallbaum (1991) stellte fest, dass durch uneinheitliche Verwahrung der Unterlagen die Daten heute nicht mehr vollständig vorliegen. Es ist in Sachsen, ebenso wie in anderen Bundesländern, jedoch zwischenzeitlich gelungen, diese Daten wieder zusammenzuführen. Von der Oberfinanzdirektion Chemnitz wird geschätzt, dass ca. 90 % der Daten wieder zugänglich sind (Kern, 2000). Daher stellt die Bodenschätzung für Deutschland die einzige annähernd flächendeckende, einheitliche und großmaßstäbige bodenkundliche Kartierung dar (Hierold, 1992; Reiche und Schleuß, 1992; Wittmann, 1992; Rothaar u. Schröder, 1987).

2.3.2 Bodenschätzung im Projektgebiet

Die in den Kapiteln 2.3.2 bis 2.3.6 geschilderte Datengrundlage stellt die Ausgangsgrundlage für die in dieser Schrift beschriebene Schätzung der Korngrößenfraktionsanteile mit Hilfe von CART dar. Kapitel 2.3.7 ist eine Kurzbeschreibung eines klassischen Ansatzes (‚Weiblen Verfahren‘), dessen Ergebnisse in Tab. 2.3. dargestellt sind. Diese Ergebnisse werden in Kap. 5 mit den Ergebnissen der *regression trees* verglichen. Die Ergebnisse der in den Kapiteln 2.3.2 bis 2.3.7 beschriebenen Abläufe

und Verfahren wurden im Rahmen einer Diplomarbeit von Weiblen (1997) erzielt. Eine detaillierte Darstellung der hier nur kurz skizzierten Vorgehensweise ist Weiblen (1997) zu entnehmen.

Das Projektgebiet befindet sich im Bereich von fünf verschiedenen Gemeindegemarkungen, die in vier Zeiträumen von drei amtlichen Bodenschätzern kartiert wurden (Weiblen, 1997). Das Datenmaterial setzt sich zusammen aus 20 einzelnen Schätzungskarten im Maßstab 1:1500 und 1.252 detaillierten Einzelflächenbeschreibungen im Schätzungsbuch. Für die flächige Ableitung der Bodenart aus den Daten der Bodenschätzung sind sowohl die Schätzungsbücher als auch die Schätzungskarten notwendig. Die Bodenschätzungskarten wurden im Rahmen einer vom Autor betreuten Diplomarbeit (Weiblen, 1997) zunächst mit dem geographischen Informationssystem ARC/INFO digitalisiert, im Anschluss daran wurde die Koordinatentransformation in das Gauß-Krüger-System durchgeführt und die einzelnen Kartenteile zusammengefügt. Die Daten des Schätzungsbuches wurden mit dem Datenbanksystem Microsoft ACCESS erfasst. Dabei wurden alle im Schätzungsbuch vermerkten Daten miterhoben. Insgesamt wurden ca. 2.300 Grablochbeschriebe mit jeweils über 60 Datenfeldern digital aufgenommen. Insgesamt kommen im Untersuchungsgebiet 47 verschiedene Bodenschätzungsbodenarten vor (siehe Tab. 2.3). Im Anschluss an die Datenerfassung wurde mit Hilfe des Geografischen Informationssystems ARC/INFO eine 1:1 Verknüpfung zwischen den Klassenflächen und den punktförmigen Grablochbeschrieben des Schätzungsbuches hergestellt, so dass für jede der 1.252 Flächen der Bodenschätzung genau ein Datensatz in der Datenbank vorgehalten wird.

2.3.3 Kartierungen nach Bodenkundlicher Kartieranleitung im Projektgebiet

Im Projektgebiet wurden von verschiedenen Institutionen Kartierungen bzw. Profilansprachen gemäß der Bodenkundlichen Kartieranleitung (nachfolgend mit BKA abgekürzt; Arbeitsgruppe Bodenkunde, 1982) durchgeführt (Weiblen, 1997). Nachfolgend genannte Datenquellen konnten für die Bildung der Übersetzungsfunktion verwendet werden.

Im Rahmen der landesweiten bodenkundlichen Kartierung im Maßstab 1:25.000 durch das damalige Geologische Landesamt Baden-Württemberg (GLA B.-W.) wurde das Arbeitsgebiet innerhalb des Blattes „6818 - Kraichtal“ kartiert (Geologisches Landesamt Baden-Württemberg, 1993). Bei dieser Kartierung wurden innerhalb des Untersuchungsgebietes insgesamt 60 Bohrstockbeprobungen mit horizontbezogener Bodenartansprache nach BKA durchgeführt. Die Bodenartbestimmung wurde durch Fingerprobe ermittelt. Die Datengrundlage wurde in Form der Flurkarte und den Aufnahmeblättern zu Verfügung gestellt und anschließend digitalisiert.

Im Rahmen des Teilprojektes 10¹ des Weiherbachprojektes wurden von der LUFA Augustenberg an 52 Stellen Texturdaten für die Tiefen von 0-30 cm und 30-60 cm analysiert und in der Projektdatenbank hinterlegt (Weiblen, 1997). Bei der Analyse wurden *Korngrößenfraktionsanteile* (Prozentangaben von Schluff, Sand und Ton) bestimmt.

Von Teilprojekt 6 wurden an 52 Stellen im Projektgebiet Bodenproben entnommen und die Korngrößenfraktionsanteile bestimmt. Die Probenahmen erfolgten ausschließlich vom Oberboden in den oberen 5 cm (Larsen, 1994). Aufgrund der starken Durchmischung des Ap-Materials durch intensives Pflügen wurden diese Beprobungen als repräsentativ für den Ap-Horizont von 0-30 cm betrachtet. Eine Übersicht der Kartierungen nach BKA im Untersuchungsgebiet ist in Tab. 2.2 abgedruckt. Die

¹ Die Teilprojekte des Weiherbachprojektes sind in Anhang 1 aufgelistet

Kartierung von Teilprojekt 17 wurde im Zeitraum April - Juni 1995 durchgeführt (Weiblen, 1997). Dabei wurde versucht, alle geologischen Einheiten, Nutzungsarten (Acker, Weinberg, Obstanlagen und Wald) sowie alle bedeutenden Reliefpositionen zu erfassen. Insgesamt wurde dabei an 100 Punkten im Untersuchungsgebiet eine Profilansprache durchgeführt. Diese erfolgten zumeist an Hand von Bohrstockbeprobungen, in wenigen Fällen wurden zur Leitprofilbestimmung eine Profilgrube ange-

Tabelle 2.2: Übersicht der Kartierungen nach Bodenkundlicher Kartieranleitung (3. Auflage) im Untersuchungsgebiet

Quelle	Profilanzahl	Horizontanzahl	Bestimmungsverfahren der Bodenart
Teilprojekt 10	52	104	Laboranalysen
GLA B.-W.	60	115	Fingeransprache
Teilprojekt 6	52	52	Laboranalysen
Teilprojekt 17	100	170	Fingeransprache

legt. Die Bodenart wurde ausschließlich durch Fingeransprache bestimmt. Die Profilaufnahmen erfolgten vor allem im südlichen Weiherbachgebiet auf der Gemarkung Gochsheim. Da zum Zeitpunkt der Kartierung die Neuauflage der

Kartieranleitung unmittelbar bevorstand, wurde die Kartierung nach dem neuen System der 4. Auflage der Bodenkundlichen Kartieranleitung (Arbeitsgruppe Bodenkunde, 1994) durchgeführt, um mit dem zukünftigen System konform zu sein. Im Laufe der Kartierung wurde aber festgestellt, dass bei Verwendung des neuen Bestimmungsschlüssels höhere Tongehalte ermittelt werden, als tatsächlich vorhanden sind. Diese Erkenntnis wurde von anderer Seite bestätigt (Weiblen, 1997). Deshalb wurden die Ergebnisse in das System der 3. Auflage transformiert. Nähere Erläuterungen hierzu finden sich in Weiblen (1997). Die zur Transformation verwendete Tabelle ist im Anhang 6 dargestellt.

Die Profilbeschreibungen und Ortskoordinaten der bodenkundlichen Kartierungen wurden ebenfalls mit EDV erfasst, so dass die Profilbeschreibungen im Datenbankformat von MS ACCESS vorliegen und eine Karte mit den Orten der Profilansprachen in dem geographischen Informationssystem ARC/INFO existiert.

2.3.4 Korngrößenfraktionsanteile

Zur besseren Handhabung der Daten der Bodenkundlichen Kartieranleitung wurden die nominalskalierten Bodenarten der BKA – sofern nicht bereits vorliegend (siehe Tabelle 2.2) – in *Korngrößenfraktionsanteile* transformiert. Korngrößenfraktionsanteile bezeichnen die Gemischanteile der Korngrößen Ton (T), Schluff (U) und Sand (S) der jeweiligen Bodenart. Die Einheit der Gemischanteile ist Prozent, bezogen auf die Masse Feinboden ohne Humusanteil, also $T + U + S = 100$. Die Transformationsfunktion einer BKA-Bodenart ergibt sich aus den Korngrößenfraktionsanteilen des Schwerpunktes der jeweiligen Fläche im Korngrößendreieck der BKA.

2.3.5 Relationen Bodenschätzung – Bodenkundliche Kartieranleitung

Nach der EDV-Aufnahme der Daten wurden die beiden Layer der Bodenschätzungsdaten und der bodenkundlichen Kartierung in ARC/INFO verschnitten, um diejenigen Polygone der Bodenschätzung zu ermitteln, in denen eine bodenkundliche Profilaufnahme durchgeführt worden war. Im nächsten Schritt wurden die entsprechenden Grablochbeschriebe der Bodenschätzung und die Profibeschreibungen der Kartierung zusammengeführt. Danach wurden - abhängig von den Tiefenangaben der Pro-

filbeschreibungen – sich entsprechende Horizontpaare (Relationen) Bodenschätzung – bodenkundliche Kartierung ausgegliedert. Insgesamt wurden 441 Horizontpaare gebildet. Die Menge der 441 Horizontpaare kann als $M:N$ oder Viele-zu-Viele-Beziehung verstanden werden. Das heißt, jeder Bodenschätzungsbodenart können mehrere BKA-Bodenarten zugeordnet sein und umgekehrt. Für eine weitere Nutzung der Daten ist jedoch eine eindeutige Beziehung im Sinne einer $N:1$ Verknüpfung erforderlich.

2.3.6 Bildung einer Übersetzungsfunktion

Für die rechtseindeutige Abbildung aus der Menge der Bodenarten nach Bodenschätzung im Untersuchungsgebiet in die Menge der Korngrößenfraktionsanteile der Bodenarten nach Bodenkundlicher Kartieranleitung wird in dieser Arbeit der Begriff *Übersetzungsfunktion* (In der Literatur auch als ‚Übersetzungsschlüssel‘ bezeichnet) verwendet. Mit Hilfe der Übersetzungsfunktion ist die flächige Übertragung der Kartierung der Bodenschätzung in das System der bodenkundlichen Kartieranleitung möglich. Bei diesem Vorgehen wird vorausgesetzt, dass das bestimmende Grabloch der Bodenschätzung für die gesamte Klassenfläche charakteristisch ist.

Im Rahmen dieser Arbeit stellt die Tabelle der 441 Horizontpaare die Grundlage für die Bildung von zwei Übersetzungsfunktionen dar. Die Bildung der ersten, nicht auf *classification and regression trees* beruhenden Übersetzungsfunktion (Weiblen, 1997; in dieser Arbeit als ‚Weiblen Verfahren‘ bezeichnet) wird im folgenden Kapitel nur kurz skizziert um den prinzipiellen Unterschied zur zweiten Übersetzungsfunktion (Ableitung mit Hilfe des CART-Verfahrens, siehe Kapitel 3.4) zu verdeutlichen. Die Ergebnisse beider Verfahren werden im Kapitel 4 mit Hilfe der in Kapitel 3.2 vorgestellten Gütekriterien verglichen.

2.3.7 Übersetzungsfunktion nach Weiblen (Weiblen-Verfahren)

Eine einfache Möglichkeit zur Bildung einer Übersetzungsfunktion besteht darin, die Korngrößenfraktionsanteile zu berechnen, falls für die betreffende Bodenschätzungsbodenart entsprechende Korngrößenfraktionsanteile nach BKA-Kartierung vorliegen, oder andernfalls die Korngrößenfraktionsanteile basierend auf einer Reihung der Bodenschätzungsbodenarten zu schätzen.

Am Anfang der Ermittlung der Übersetzungsfunktion nach Weiblen wurde der Datensatz der 441 Horizontpaare durch die Variable Bodenschätzungsbodenart gruppiert und nach zunehmendem Tongehalt geordnet. Falls in einer Bodenschätzungsbodenartgruppe mehr als fünf BKA-Bodenarten vorlagen, wurden die Korngrößenfraktionsanteile durch die Medianwerte der BKA-Bodenarten berechnet. Dieses Verfahren konnte bei 17 Bodenschätzungsbodenarten durchgeführt werden. Bei 24 von 47 im Untersuchungsgebiet vorkommenden Bodenschätzungsbodenarten konnten keine Horizontpaare gebildet werden d. h. in den betreffenden Klassenflächen der Bodenschätzung fand keine Beprobung nach der bodenkundlichen Kartieranleitung statt. In diesen und in weiteren fünf Fällen, bei denen weniger als fünf BKA-Bodenarten vorlagen, wurden die einzelnen Korngrößenfraktionanteile linear interpoliert. Bei den restlichen 6 Bodenschätzungsbodenarten konnte nicht interpoliert werden, da keine BKA-Bodenarten mit höherem Tongehalt vorlagen. In diesen Fällen wurden die Korngrößenfraktionsanteile mittels einer Regressionsgeraden der 15 berechneten Bodenschätzungsbodenarten extrapoliert. Die durch Berechnung oder Schätzung bestimmten Korngrößenfraktionsanteile wurden anschließend nor-

miert um eine Summe der Gehalte von eins zu gewährleisten. Danach wurde mittels des Korngrößendreiecks die Bodenartbezeichnung nach der bodenkundlichen Kartieranleitung bestimmt.

Das Ergebnis des oben beschriebenen Verfahrens der Bildung der Übersetzungsfunktion ist in Tab. 2.3 dargestellt, die Tabelle der 441 Wertepaare befindet sich auf der beigelegten CD-ROM (,Relationen der Daten der Bodenschätzung vs. Bodenkundliche Kartieranleitung.xls'). Aus den Spalten 3 und 4 der Tab. 2.3 geht hervor, dass die Gruppen, bei denen eine Berechnung der Korngrößenfraktionsanteile möglich war, 94,8 % der 3590 im Schätzungsbuch enthaltenen Bodenhorizonte umfassen. Die Interpolationsmethode brauchte bei nur 3,4 % und die Extrapolationsmethode bei nur ca. 1,9 % der Flächen angewendet zu werden um die Korngrößenfraktionsanteile zu bestimmen.

Tabelle 2.3: Übersetzungsfunktion der Bodenarten nach Weiblen (1997)

Bodenartbezeichnung im Feldbuch der Bodenschätzung (BS)	Bodenartbezeichnung in der Karte der Bodenschätzung	Anzahl der Horizonte im Gebiet	Anzahl der Wertepaare BS-BKA im Gebiet	Verfahren 1 = Berechnung 2 = Interpolation 3 = Extrapolation	Geschätzter Tongehalt [%]	Geschätzter Schluffgehalt [%]	Geschätzter Sandgehalt [%]	Bodenart nach BKA
SI3-SI4	LS	1	1	1 ²	2.5	5.0	92.5	S
F3SI3	IS	113	6	1	10.1	79.4	10.5	UI2
SI4	SL	2	2	1	2.5	5.0	92.5	S
f3SI3-f3SI4	SL	181	33	1	10.1	79.4	10.5	UI2
Lfs4-f3SI3	SL	91	5	1	3.7	88.2	8.1	U
f3SI4	SL	25	13	1	9.9	77.5	12.6	UI2
f3SI4-(Lfs4)	SL	15	0	2	12.0	76.7	11.3	UI3
Lfs4-f3SI4	sL	206	35	1	14.1	75.8	10.1	UI3
Ls4	sL	1	0	2	13.1	76.7	10.2	UI3
Lfs4-(f3SI4)	sL	8	0	2	12.1	77.6	10.3	UI3
Lfs4	sL	367	22	1	11.0	78.5	10.4	UI2
Lfs3-f3SI4	sL	5	2	2	11.9	77.7	10.4	UI2
Lfs4-(Lfs3)	sL	1	0	2	12.8	76.8	10.4	UI3
Ls3	sL	12	0	2	13.6	76.0	10.3	UI3
Lfs3-Lfs4	sL	195	28	1	14.5	75.2	10.3	UI3
Ls2-Ls3	sL	1	0	2	15.3	75.2	9.5	UI3
Lfs4-Lschlu3	sL	1	0	2	16.1	75.2	8.7	UI3
Lfs3-(Lfs4)	L	19	0	2	16.9	75.3	7.8	UI3
Lfs3	L	769	111	1	17.7	75.3	7.0	UI4
Lfs3-(Lfs2)	L	5	0	2	19.2	73.8	7.0	UI4
Lfs2-Lfs4	L	2	1	2	20.6	72.4	7.0	UI4
Lfs2-Ls3	L	1	1	2	22.1	70.9	7.0	UI4
Lfs2-Lfs3	L	101	19	1	23.5	69.5	7.0	Lu
Lfs3-Lschlu3	L	3	0	2	23.1	69.5	7.4	Lu
Ls3-Lkr3	L	1	0	2	22.7	69.6	7.7	Lu
Lfs3-Lkr3	L	1	1	2	22.3	69.6	8.1	Lu
Lschlu3-Lschli3	L	1	0	2	21.9	69.7	8.5	Lu
Lkr3-Ls2	L	1	0	2	21.4	69.7	8.8	Lu
Lfs2-(Lfs3)	L	4	0	2	21.0	69.7	9.2	Lu
Lfs2	L	938	97	1	20.6	69.8	9.6	Lu
Lfs2-(Lkr3)	L	17	0	2	23.7	64.9	11.4	Lu
Ls3-Lt3	L	4	6	1	53.8	32.1	14.1	Tl
Lfs2-Lkr3	L	153	14	1	26.8	60.0	13.2	Lu
Lfs1	L	5	0	2	33.8	52.1	14.1	Ltu
Lkr3-(Lfs2)	L	6	0	2	40.9	44.1	15.0	Lt3
Lkr3	L	181	17	1	47.9	36.2	15.9	Tl
Lkr3-(Lt2)	L	2	0	2	47.9	36.2	15.9	Tl
Lkr3-Lschw3	L-LT	48	13	1	47.9	36.2	15.9	Tl
Lschw3-(Lkr3)	LT	8	0	2	50.9	34.1	15.0	Tl
Lschw3	LT	28	10	1	53.8	32.1	14.1	Tl
Lkr3-Lt3	LT	1	0	3	55.0	30.7	14.3	Tl
Lschw4	LT	1	0	3	56.2	29.3	14.5	Tl
Lschw3-Lt3	LT	44	4	3	57.3	28.0	14.7	Tl
Lschw3-Lt3-Tl3	LT	2	0	3	58.4	26.6	15.0	Tl
Lt3	LT	17	0	3	59.6	25.2	15.2	Tl
Tl3-T	T	2	0	3	60.7	23.8	15.5	Tl

² Diese Bodenarten nehmen aufgrund ihrer geologischen Besonderheit (vernachlässigbares Schilfsandsteinvorkommen, siehe Kap. 2.1.2) eine Sonderstellung im Untersuchungsgebiet ein; sie wurden daher mit dem Berechnungsverfahren bestimmt

3 Methoden

3.1 Grundzüge robuster Statistik

Robustheit bezeichnet die Unempfindlichkeit gegenüber kleinen Änderungen der Modellannahmen (Huber, 1981). Traditionell wurden robuste Verfahren in der Mathematik im Zusammenhang mit nichtparametrischen und verteilungsfreien Verfahren genannt. Robuste Verfahren können jedoch nicht mit nichtparametrischen oder verteilungsfreien Verfahren gleichgesetzt werden, da robuste Methoden vorwiegend bei parametrischen Verteilungsmodellen angewendet werden und nichtparametrische und verteilungsfreie Verfahren z. T. wenig robust sind (Huber, 1981). Robuste Methoden sind daher parametrischen Ansätzen viel näher als nichtparametrischen und verteilungsfreien, sie unterscheiden sich von nichtrobusten, gewöhnlichen Verfahren insofern, als bei robusten Verfahren die Voraussetzungen parametrischer Verfahren nicht als stillschweigend erfüllt betrachtet werden. Die möglichen Abweichungen gegenüber den Voraussetzungen parametrischer Verfahren werden bei robusten Verfahren formal berücksichtigt (Huber, 1981).

Ein häufig gegenüber robuster Statistik vorgebrachter Einwand ist, dass robuste Statistik nicht notwendig sei, da eine zweistufige Datenanalyse mit einer sinnvoll durchgeführten Ausreißerelimination der robusten Statistik ebenbürtig sei. Dabei wird im ersten Schritt der Datensatz durch Anwendung einer Ausreißerregel bereinigt und im zweiten Schritt werden klassische Schätz- und Testverfahren auf die verbliebenen Daten angewendet. Es gibt jedoch drei Gründe, die gegen ein solches Datenmanagement sprechen (Huber, 1981):

1. Bei mehrdimensionalen Regressions-Problemen sind Ausreißer sehr schwer zu erkennen (Hampel et al., 1986).
2. Selbst wenn der vorliegende Datensatz normalverteilt ist und nur einige größere Fehler 'eingestreut' sind, sind die bereinigten Daten nicht normalverteilt. Die Situation ist noch wesentlich kritischer zu beurteilen, wenn keine Normalverteilung vorliegt, der Datensatz fehlerbehaftet ist, eine Ausreißerelimination durchgeführt wurde und dann eine Normalverteilung angenommen wird. Die Leistungsfähigkeit einer derartigen Vorgehensweise ist weitaus schwieriger zu bestimmen, als die einer robusten Methode. Die klassische Theorie der Normalverteilung ist daher auf *bereinigte* Daten nicht anwendbar.
3. Empirische Untersuchungen (Hampel, 1986) ergaben, dass die besten Ausreißerregeln nicht so leistungsstark sind, wie die besten robusten Prozeduren. Robuste Methoden weisen einen fließenden Übergang zwischen Ablehnung und Verbleib einer Beobachtung im Datenensemble auf und sind deshalb den Ausreißerregeln überlegen. Der fließende Übergang zwischen Ablehnung und Verbleib wird im Unterkapitel 'Einflussfunktion' näher erläutert.

3.1.1 Robuste Kenngrößen

Die wichtigsten quantitativen Kenngrößen, welche die Robustheitseigenschaften einer Schätzfunktion T_N beschreiben, sind die *Sensitivitätskurve*, die *Einflussfunktion*, die *gross error sensitivity*, die *local shift sensitivity*, der *rejection point* und der *Bruchpunkt*, sie werden in Abb. 3.1 beispielhaft für die Schätzfunktionen Mittelwert und Median dargestellt.

3.1.1.1 Sensitivitätskurve

Die *Sensitivitätskurve* SC (engl. *sensitivity curve*) einer Schätzfunktion beschreibt die Veränderung einer Schätzfunktion, wenn zu einer Stichprobe x_1, \dots, x_{N-1} noch ein Wert x hinzugenommen wird (Tukey, 1970):

$$SC(x; x_1, \dots, x_{N-1}; T_N) = N [T_N(x_1, \dots, x_{N-1}, x) - T_{N-1}(x_1, \dots, x_{N-1})] \quad (3.1)$$

Die Sensitivitätskurve ist zur allgemeinen Beschreibung von Robustheitseigenschaften einer Schätzfunktion nur bedingt geeignet, da sie außer von der Schätzfunktion selbst auch von den konkreten Beobachtungswerten und deren Anzahl abhängig ist.

3.1.1.2 Einflussfunktion

Die *Einflussfunktion* IF (*influence function*; gelegentlich wird im deutschen Sprachgebrauch auch die ältere [und weniger exakte] Bezeichnung *Einflusskurve* verwendet) ist ein Grenzwert der Sensitivitätskurve bezüglich Stichprobenwerten aus einer Verteilung F (Hampel, 1968, 1971, 1974):

$$IF(x; F; T) = \lim_{N \rightarrow \infty} SC(x; x_1, \dots, x_{N-1}; T_N) \quad (3.2)$$

Die Einflussfunktion beschreibt die Veränderung einer Statistik auf einen zusätzlichen Stichprobenwert, wenn von einer sehr großen Stichprobenzahl aus einer Verteilung F ausgegangen wird. Die Einflussfunktion ist zur allgemeinen Charakterisierung der Robustheit einer Stichprobenfunktion gut geeignet, da sie unabhängig ist von konkreten Beobachtungswerten und deren Anzahl. Die Einflussfunktion stellt ein wichtiges Werkzeug zur Beurteilung von Robustheitseigenschaften dar; von ihr lassen sich weitere wichtige Größen ableiten, z. B. wie die *gross-error sensitivity*, *local-shift sensitivity* und der *rejection point*.

3.1.1.3 Gross-Error sensitivity

Die *gross-error sensitivity* $\gamma^*(T, F)$ ist das Supremum des absoluten Betrages der Einflussfunktion (Hampel, 1968, 1974):

$$\gamma^*(T, F) = \sup_x |IF(x; T; F)| \quad (3.3)$$

Die *gross-error sensitivity* ist ein Maß für den größten Einfluss, den ein einzelner Wert auf eine Schätzfunktion haben kann. Daher ist es erstrebenswert, dass $\gamma^*(T, F)$ endlich ist, in diesem Falle heißt es, T ist B -robust auf F (Rousseeuw, 1981).

3.1.1.4 Local-Shift sensitivity

Die *local-shift sensitivity* λ^* (Hampel, 1968, 1974) beschreibt die größte Änderung der Schätzfunktion, welche durch eine kleine Änderung einer Beobachtung hervorgerufen werden kann. Kleine ‘Wackler’ von Stichprobenwerten können durch Runden, Gruppieren oder andere kleine Ungenauigkeiten bei der Beobachtung verursacht werden. Diese kleinen Verschiebungen von einem Wert x zum Wert y können mathematisch als Hinzunahme einer neuen Beobachtung y und Streichen der Beobachtung x interpretiert werden. Die sich daraus ergebende maximale Änderung der Schätzfunktion kann als Supremum der normierten Differenz der Einflussfunktionen y und x beschrieben werden (Hampel, 1968, 1974):

$$\lambda^* = \sup_{x \neq y} \frac{|IF(y; T; F) - IF(x; T; F)|}{|y - x|} \quad (3.4)$$

Ein unendlicher Wert von λ^* kann allerdings aufgrund der Normierung auch bei nur begrenzten Änderungen der Schätzfunktion auftreten.

3.1.1.5 Rejection Point

Das Vermögen, Ausreißer vollständig zu vernachlässigen, liegt vor, wenn die Einflussfunktion außerhalb eines endlichen Bereiches den Wert Null annimmt. Liegen Beobachtungen mit Messwerten vor, die außerhalb dieses Bereiches liegen, haben diese keinerlei Einfluss auf die Schätzfunktion. Bei Vorliegen einer symmetrischen Wahrscheinlichkeitsdichte mit Mittelwert Null wird der *rejection point* wie folgt definiert (Hampel, 1968, 1974):

$$\rho^* = \inf \left\{ r > 0 \mid IF(x; T; F) = 0 \text{ und } |x| > r \right\} \quad (3.5)$$

Existiert kein derartiges r , ist gemäß der Definition des Infimums $\rho^* = \infty$.

3.1.1.6 Breakdown Point

Die allgemeine Definition des *breakdown points* ε^* (Bruchpunkt; Hampel, 1968), ist in der praktischen Arbeit schwer handhabbar, so dass hier eine vereinfachte Definition vorgestellt wird. Der Bruchpunkt ε_N^* der Stichprobe (x_1, \dots, x_N) einer Schätzfunktion T_N ist definiert als (Hampel et al., 1986)

$$\varepsilon_N^*(T_N; x_1, \dots, x_N) = \frac{1}{N} \max \left\{ m \mid \sup_{y_1, \dots, y_m} |T_N(z_1, \dots, z_N)| < \infty \right\} \quad (3.6)$$

wobei die Stichprobe (z_1, \dots, z_N) aus der Stichprobe (x_1, \dots, x_N) durch Ersetzen von M Beobachtungen x_{i_1}, \dots, x_{i_M} , $1 \leq i, j \leq N$, durch (y_1, \dots, y_M) hervorgeht. In vielen Fällen kann mittels Grenzwertbildung $\lim_{N \rightarrow \infty} \varepsilon_N^*$ der *allgemeine Bruchpunkt* ε^* gewonnen werden. Der Bruchpunkt einer Stichprobe kann als der maximale Anteil von Ausreißern in einer Stichprobe verstanden werden, ohne dass sich die Schätzfunktion unbeschränkt ändert. Der Bruchpunkt hat immer einen Wert kleiner oder gleich 0.5 (Hampel et al., 1986).

3.1.1.7 Beispiel

Die oben dargestellten Robustheitseigenschaften werden am Beispiel der ‚klassischen‘ Lageparameter Mittelwert und Median verdeutlicht. Der Mittelwert ist besonders empfindlich, wenn Beobachtungen mit großen Fehlern behaftet sind, was auf den Median jedoch keine Auswirkungen hat, die *gross-error sensitivity* ist daher beim Mittelwert unbeschränkt, beim Median hingegen beschränkt. Das Verhalten der beiden Lageparameter gegenüber kleineren Fehlern („Wackler“) ist jedoch genau umgekehrt. Der Mittelwert reagiert schwach auf kleinere Messwertschwankungen (beschränkte *local shift sensitivity*), der Median weist eine unbeschränkte *local shift sensitivity* auf, da im Bereich des Zentralwertes schon eine einzige kleine Messwertschwankung zu einer sprunghaften Änderung des Lageparameters führen kann. Diese Eigenschaften dokumentieren sich beim maximalen Sprung der Einflussfunktion. Beim Mittelwert kann sich bereits ein einzelner großer Fehler beliebig stark auf den Schätzwert auswirken, daher ist der Bruchpunkt einer Stichprobe 0, beim Median hingegen kann der Anteil an Ausreißern bis zu 50 % betragen, bevor diese sich überhaupt erst auf den zu schätzenden Wert auswirken. Beide Lageparameter sind außerdem nicht in der Lage, Messwerte außerhalb eines bestimmten Bereiches (Ausreißer) prinzipiell zu ignorieren, ihr *rejection point* ist daher unendlich. Mittelwert und Median haben daher sowohl einige positive als auch negative Robustheitseigenschaften und sind demnach als wenig robust zu charakterisieren.

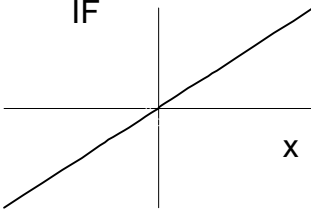
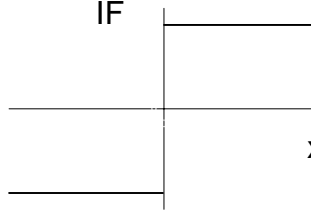
Eigenschaft	Arithmetisches Mittel	Median
Schematisierte Einflussfunktion		
<i>gross-error sensitivity</i>	unbeschränkt	beschränkt
<i>local-shift sensitivity</i>	Beschränkt	unbeschränkt
Maximaler Sprung der IF bei Normalverteilung	0	$\sigma\sqrt{2\pi}$
<i>Rejection point</i>	∞	∞
Bruchpunkt einer Stichprobe	0	0.5

Abb. 3.1: Vergleich der Robustheitseigenschaften des Mittelwertes und des Median

3.1.2 M - Schätzer

Der Name M-Schätzer (Huber, 1964) stammt von ‘generalized maximum likelihood’. Ein (skalenabhängiger) M-Schätzer T_N für die Lokation ergibt sich generell aus der Lösung der Minimierungsaufgabe

$$\sum_{i=1}^N \rho(x_i - T_N) = \min \quad (3.7)$$

wobei ρ eine beliebige Funktion ist (Huber, 1981).

Setzt man $\psi(x - \theta) = \frac{\partial}{\partial \theta} \rho(x - \theta)$, so wird anstelle von Gl. 3.7 die Gleichung

$$\sum_{i=1}^N \psi(x_i - T_N) = 0 \quad (3.8)$$

gelöst.

Normalerweise sind M-Schätzer von der Wahl der Messskala abhängig. Eine Ausnahme bilden der Median und der Mittelwert, die Sonderfälle der M-Schätzer und skaleninvariant sind. Skalenabhängige M-Schätzer können skaleninvariant gemacht werden, in dem bei der Minimierungsaufgabe eine Skalenschätzung S_N mit einbezogen wird. Zur Skalenschätzung bietet sich der Median der Medianabweichung, MAD, an (Hampel et al., 1986), der mit Hilfe des Faktors c so standardisiert wird, dass bei der angenommenen Modellverteilung Konsistenz vorliegt (Huber, 1981):

$$S_N = c \cdot \text{MAD}_N = c \cdot \text{med} \{x_i - \text{med} \{x_i\}\} \quad (3.9)$$

Bei Normalverteilung ist $c = 1.483$ (Hampel et al., 1986)

Mit dieser Skalenschätzung kann die Lageschätzung durch Lösen der modifizierten Minimierungsaufgabe

$$\sum_{i=1}^N \rho\left(\frac{x_i - T_N}{S_N}\right) = \min \quad (3.10)$$

skaleninvariant gemacht werden. Von den Sonderfällen Median und Mittelwert abgesehen, erfordert die Berechnung von M-Schätzern immer den Einsatz numerischer Verfahren zur Lösung der Gleichung 3.8, z. B. des Newton-Raphson-Verfahrens.

3.1.2.1 Huber-k-Schätzer

Die *Huber-k-Schätzer* stellen eine Klasse von M-Schätzern dar, bei denen die Funktionen des Mittelwertschätzers und des Medianschätzers abschnittsweise kombiniert sind. Diese Klasse enthält den Mittelwert- und Medianschätzer als Spezialfälle, vereint aber die Vorteile der beiden Schätzer. Huber-k-Schätzer sind durch folgende Definitionen für ρ oder ψ festgelegt (Huber, 1964; Hartung, 1986):

$$\rho(x) = \begin{cases} x^2 / 2 & \text{falls } |x| \leq k \\ k \cdot |x| - k^2 / 2 & \text{falls } |x| > k \end{cases} \quad \text{bzw.} \quad (3.11)$$

$$\psi(x) = \begin{cases} x & \text{falls } |x| \leq k \\ k \cdot \text{sgn}(x) & \text{falls } |x| > k \end{cases} \quad (3.12)$$

wobei k eine positive, reelle, Konstante, den sog. Cut-off-Punkt, darstellt. Meist wird für k die Medianabweichung (Hampel, 1974) verwendet.

In Abb. 3.2 ist der qualitative Verlauf einer ψ – Funktion dargestellt, welche einen Huber-k-Schätzer

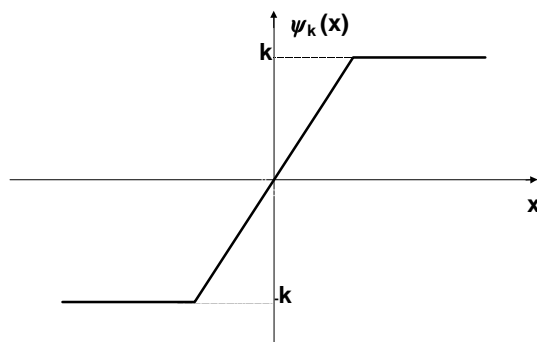


Abb. 3.2 Huber-k-Schätzer definierende ψ – Funktion mit Cut-Off-Punkt k
(Erläuterung im Text)

für eine Verteilung mit Erwartungswert Null definiert. Aufgrund der Eigenschaften der Huber-k-Schätzer ist die den Schätzer definierende ψ – Funktion proportional zur Einflussfunktion des Schätzers. Es wird deutlich, dass Huber-k-Schätzer hinsichtlich der Einflussfunktion eine Synthese zwischen Median und Mittelwert darstellen und die Vorteile beider Verfahren kombinieren: Die Einflussfunktion ist beschränkt, weist keine Sprungstellen auf und ist punktsymmetrisch, daher sind sowohl *local-shift sensitivity* als auch *gross-error sensitivity* beschränkt; der Schätzer ist somit qualitativ (Definition siehe Anhang) und B(ias)-robust mit

einem Bruchpunkt von 0.5 (Hampel et al., 1986). Für weitere Schätzer zur Bestimmung der Lokation, wie A-, D-, P-, S- und W-Schätzer wird auf Lax (1975), Wolfowitz (1957), Johns (1979), Rousseeuw & Yohai (1984) bzw. Tukey (1970) verwiesen. Eine Übersicht ist in Hampel et al. (1986) zu finden.

3.2 Gütekriterien

In diesem Kapitel wird ein Instrumentarium eingeführt, mit dem die Güte räumlicher Schätzungen der beiden Fallbeispiele in Kapitel 4 und 5 charakterisiert wird. Es wird davon ausgegangen, dass der zur Verfügung stehende Datensatz mit kleinen und/oder großen Fehlern behaftet sein kann, auf deren Ursache im vorigen Kapitel eingegangen wurde. Gesucht sind daher Verfahren zur Gütebeschreibung, welche die Existenz von derartigen Fehlern im Datensatz nicht a priori negieren, sondern von fehlerbehafteten Daten ausgehen und dies formal zu berücksichtigen versuchen. Bei nichtrobusten Kriterien kann bereits ein einzelner Messwertausreißer das Gütemaß über alle Grenzen beeinflussen, so dass die Aussagekraft derartiger Kriterien eingeschränkt ist. Daher ergänzen robuste Kriterien den bisherigen Katalog von weitgehend nichtrobusten Gütekriterien und können dazu beitragen, Mindestanforderungen an Schätzverfahren schärfer zu formulieren. Sämtliche Gütekriterien, die eine Weiterentwicklung

der klassischen Kriterien unter Berücksichtigung von Robustheitsaspekten darstellen, sind nach Kenntnis des Autors neu.

3.2.1 Allgemeine Korrelationsschemata bei Schätzungen

In diesem Abschnitt werden die wichtigsten Schemata des Zusammenhanges von Messwerten und Schätzwerten aufgezeigt und grob klassifiziert. Diese Klassifikation ist Grundlage für die im nächsten Kapitel erfolgende Erarbeitung von Gütekriterien, die jeweils die verschiedenen Zusammenhänge charakterisieren sollen. Durch die Klassifizierung wird der Nutzen der robusten Gütekriterien deutlicher, da bei nichtrobusten Gütekriterien die verschiedenen Korrelationsschemata weniger gut herausgearbeitet werden können. Spezifische und quantifizierbare Gütemaße sind für einen objektiven Verfahrensvergleich unbedingt erforderlich, da rein qualitative Beschreibungen nur bedingt aussagekräftig sind. Die Klassifizierungsmöglichkeiten dienen letztlich dazu, Daten möglichst erschöpfend nutzen zu können, um bei gegebener Problemstellung eine breite Grundlage für die Bewertung zur Auswahl stehender Schätzverfahren zu schaffen. Bei der Anwendung von Schätzverfahren zur räumlichen Schätzung können verschiedene Typen von Abweichungen vom zu schätzenden Wert auftreten. Die wichtigsten Abweichungen sind in Abb. 3.4 schematisch dargestellt und erläutert.

3.2.2 Analyse der Momente von Messung und Schätzung

Eine Möglichkeit, die Güte einer räumlichen Schätzung zu quantifizieren, besteht darin, die Momente der gemessenen und der geschätzten Werte auf Unterschiede zu analysieren. Im folgenden bezeichnet y_i den Messwert an der Stelle i und \hat{y}_i den Schätzwert für y_i , der aus den Messwerten $y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_N$ geschätzt ist.

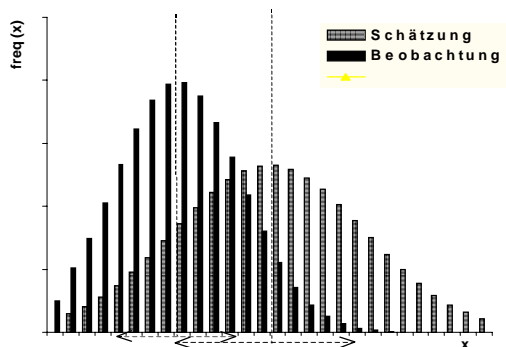


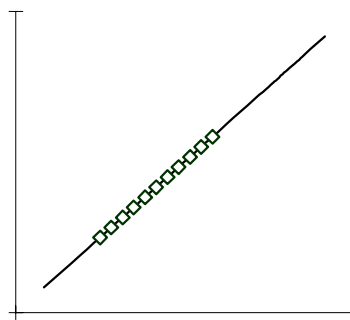
Abb. 3.3: Schematische Darstellung der Wahrscheinlichkeitsdichten von Beobachtungen und Schätzungen

Dargestellt sind die Wahrscheinlichkeitsdichten der Messwerte und der Schätzwerte. Gestrichelte Linien kennzeichnen die Lage des Mittelwertes, die Länge der Doppelpfeile gibt die Standardabweichung wieder

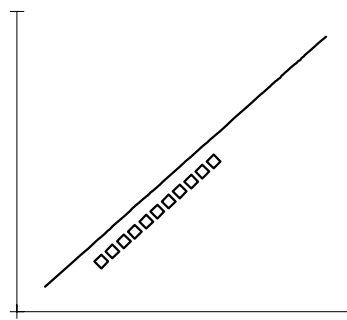
3.2.2.1 Bias

Ein wichtiger Kennwert der Wahrscheinlichkeitsdichte ist der Mittelwert. Die Differenz zwischen den Mittelwerten der Beobachtungen und der Schätzwerte heißt *Bias* der Schätzung (*average error* nach Janssen & Heuberger; 1995). Der Bias gibt Aufschluss über im Mittel zu hohe oder zu niedrige Schätzwerte:

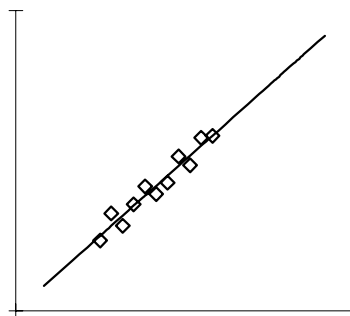
$$bias = \frac{1}{N} \left(\sum_{i=1}^N y_i - \sum_{i=1}^N \hat{y}_i \right) \quad (3.13)$$



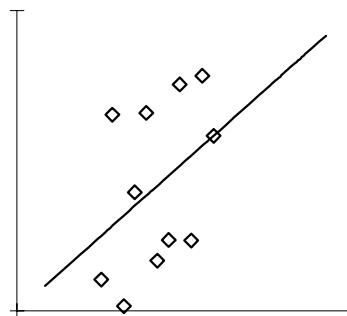
a) Idealer Zusammenhang



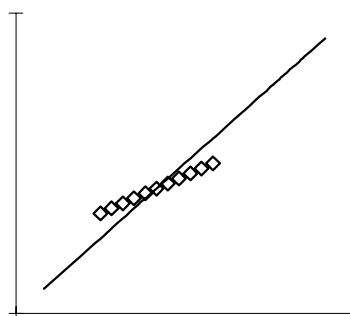
b) Schätzung verzerrt



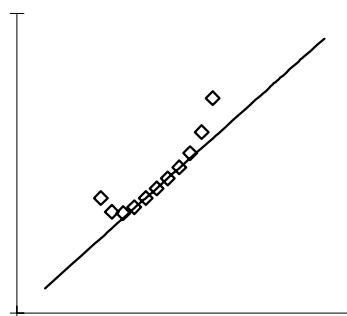
c) Kleine Schätzabweichung



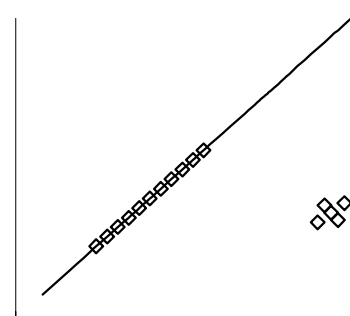
d) Schätzung unbestimmt



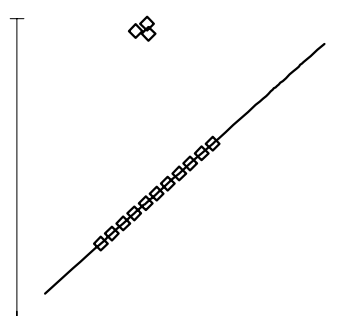
e) Systematischer Fehler



f) Nichtlineare Korrelation



g) Messwerte ausreißerbehaftet



h) Schätzwerte ausreißerbehaftet

Abb. 3.4 a) - h): Wichtige Korrelationsschemata zwischen Mess- und Schätzwerten

Dargestellt sind auf der Abszisse die gemessenen und auf der Ordinate die geschätzten Werte. Erläuterungen: a) Die erste Winkelhalbierende, die das ideale Schätzverfahren charakterisiert, ist als durchgezogene Linie gekennzeichnet;

b) verzerrte Schätzung; hier sind die Lageparameter von Schätzung und Messung verschieden;

c) Schätzung mit kleinem, zufälligem, Fehler; die Residuen zwischen Schätzwert und Messwert sind zufällig und unabhängig;

d) Unbestimmte Schätzer zeichnen sich dadurch aus, dass die Residuen größer sind als die Schätzabweichungen des besten konstanten Schätzers;

e) Schätzungen mit systematischem Fehler zeigen einen linearen Trend der Residuen, so dass die Messwerte durch die Schätzwerte unter- und/oder überschätzt werden;

f) Nichtlinearer Trend der Residuen;

g) Ausreißer bei den Messwerten;

h) Ausreißer bei den Schätzwerten.

Ein Schätzer oder Schätzverfahren heißt *biasfrei*, falls der Erwartungswert des *bias* gleich Null ist. Der Bias allein ist jedoch weder ein notwendiges noch hinreichendes Kriterium für systematische Fehler. Die Bias-Schätzung kann robuster gemacht werden, indem z. B. statt des Mittelwertes robuste Lageschätzer verwendet werden:

$${}^{robust}bias = T_N(y_1, \dots, y_N) - T_N(\hat{y}_1, \dots, \hat{y}_N) \quad (3.14)$$

T_N bezeichnet hierbei eine robuste Lageschätzung, wie in Kap. 3.1 beschrieben, etwa durch einen Huber-k-Schätzer. Die Gleichung 3.14 wurde unter mehreren möglichen gewählt, da sie dem Ziel einer einfachen Charakterisierung der Gebietseigenschaft eines Schätzverfahrens am nächsten kommt. Die robuste Biasschätzung ergänzt den Informationsgehalt der nichtrobusten Biasschätzung. Biasschätzungen sind insbesondere für den Vergleich mehrerer Schätzverfahren wichtig.

3.2.2.2 Verhältnis der Mittelwerte der Verteilungen

Ein anschauliches Gütemaß ist das Verhältnis der Mittelwerte der Verteilungen (coefficient of residual mass, *crm*; [Loague and Green, 1991]):

$$crm = \frac{\sum_{i=1}^N y_i - \sum_{i=1}^N \hat{y}_i}{\sum_{i=1}^N y_i} \text{ mit } \sum_{i=1}^N y_i \neq 0 \quad (3.15)$$

Das Maß *crm* gibt Aufschluss über die *relative* Abweichung der Mittelwerte und ist daher für die Beurteilung der Güte eines Schätzverfahrens eine wichtige Grundlage.

Eine robustere Version des Verhältnisses zwischen den Lagemäßen der Schätzwerte und den Beobachtungen mindert den Einfluss von Ausreißern und kleinen Fehlern der Messungen auf das Gütemaß. Deshalb wird eine robustere Version von *crm* vorgeschlagen:

$${}^{robust}crm = \frac{T_N(y_1, \dots, y_N) - T_N(\hat{y}_1, \dots, \hat{y}_N)}{T_N(y_1, \dots, y_N)} \text{ für } T_N(y_1, \dots, y_N) \neq 0 \quad (3.16)$$

Wenn ${}^{robust}crm$ den Wert Null annimmt, kann eine optimale Schätzung vorliegen.

3.2.2.3 Höhere Momente der Verteilungen

Weitere Kennziffern lassen sich durch den Vergleich der höheren, zentrierten Momente $m_r(y)$ der beiden Wahrscheinlichkeitsdichten ermitteln.

Das zentrierte Moment $m_r(y)$ einer Variable y lässt sich für nicht gruppierte Einzeldaten allgemein wie folgt schätzen (Ferschl, 1985; vereinfacht nach Sachs, 1997):

$$\hat{m}_r(y) = \frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^r, \quad r \in \{1, 2, \dots\}, \quad \bar{y} = \frac{1}{N} \sum_{i=1}^N y_i \quad (3.16a)$$

Als robuste Version der Potenzmomente wird vorgeschlagen:

$${}^{robust}m_r(y_i) = T_N([y_i - T_N(y_i)]^r), \quad r \in \{1, 2, \dots\} \quad (3.16b)$$

Die entsprechenden Momentvergleiche lauten:

$$cmp_r = m_r(y_i) - m_r(\hat{y}_i) \quad (3.16c)$$

bzw.

$${}^{robust}cmp_r = {}^{robust}m_r(y_i) - {}^{robust}m_r(\hat{y}_i) \quad (3.16d)$$

Als Beispiel ergibt sich im nichtrobusten Fall für $r = 2$, also der Varianz:

$$m_2(y) = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{y})^2 \quad \text{bzw.} \quad m_2(\hat{y}) = \frac{1}{N-1} \sum_{i=1}^N (\hat{y}_i - \bar{\hat{y}})^2 \quad (3.17)$$

Damit kann die Differenz der Varianzschätzungen der Schätzwerte und der Beobachtungen (difference of variances; dv) berechnet werden:

$$dv = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{y})^2 - \frac{1}{N-1} \sum_{i=1}^N (\hat{y}_i - \bar{\hat{y}})^2 \quad (3.18)$$

bzw.

$$\begin{aligned} {}^{robust}dv = & T_N([y_1 - T_N(y_1, \dots, y_N)]^2, \dots, [y_N - T_N(y_1, \dots, y_N)]^2) - \\ & T_N([\hat{y}_1 - T_N(\hat{y}_1, \dots, \hat{y}_N)]^2, \dots, [\hat{y}_N - T_N(\hat{y}_1, \dots, \hat{y}_N)]^2) \end{aligned} \quad (3.19)$$

3.2.3 Analyse der Schätzabweichungen von Messwerten und Schätzwerten

Die Qualität eines Schätzverfahrens kann auch durch verschiedene Statistiken der Schätzabweichungen (Residuen) $r_i = y_i - \hat{y}_i$ charakterisiert werden. Oben wurde bereits der *bias* als wichtigste Statistik der Schätzabweichungen vorgestellt; in den folgenden Abschnitten werden weitere Statistiken vorgestellt.

Eine wichtige Forderung ist die nach minimaler Schätzvarianz:

$$\frac{1}{N-1} \sum_{i=1}^N (y_i - \hat{y}_i)^2 = \min \quad (3.19a)$$

3.2.3.1 Maximale Schätzabweichung

Die maximale Schätzabweichung berechnet sich als Betrag der größten auftretenden Differenz zwischen den gemessenen und den geschätzten Werten; er sollte möglichst klein sein (*MaxAE* nach Janssen & Heuberger; 1995).

$$me = \max_{1 \leq i \leq N} |y_i - \hat{y}_i| \quad (3.20)$$

Die maximale Schätzabweichung erlangt besondere Bedeutung, wenn die Schätzwerte die Eingangsvariable für eine hoch nichtlineare Anwendungsfunktion (z. B. Stofftransportmodelle wie WHNSIM [Huwe, 1992] und DYNAMIT [Huwe et al. 1997]) darstellen. Dann können Schätzwerte mit besonders hohen Schätzabweichungen zu Ausreißern der Zielgröße der Anwendungsfunktion führen und deren Güte erheblich verschlechtern.

Eine Verallgemeinerung der maximalen Schätzabweichung, die Gl. 3.20 als Grenzfall enthält, stellen die Perzentile me_r (Sachs, 1997) der Residuen dar (*UppAE* für $r = 75$ nach Janssen & Heuberger; 1995):

$$me_r = |y_i - \hat{y}_i|_r; \quad 50 < r < 100 \quad (3.21)$$

Dabei ist me_r das r -te Perzentil, der Wert mit der nächsten ganzen Ordnungszahl $(N+1)r/100$ der aufsteigend geordneten Stichprobe mit N Elementen. Bei geeignet gewähltem r ist dieses Maß weniger sensitiv gegenüber Ausreißern und kann Informationen über deren Struktur liefern. Der Bruchpunkt der robusten Version me_r (Gl. 3.21) beträgt $(100-r)/100$, die maximale Schätzabweichung me (Gl. 3.20) hat dementsprechend den Bruchpunkt Null und ist deshalb nicht robust.

Die Perzentile der Residuen sind einfach zu berechnen und können Informationen über die Struktur der Residuen liefern, die bei Vergleichen von verschiedenen Schätzverfahren interessant sein können.

3.2.3.2 Wurzel der mittleren quadrierten Schätzabweichung

Ein Kriterium zur Beschreibung der Residuen ist die Wurzel der mittleren quadrierten Schätzabweichung (root mean square of residuals; *rmsr* nach Loague & Green, 1991). Es ermöglicht (in Ergänzung zur maximalen Schätzabweichung) Aussagen über die mittlere Größe der Schätzabweichungen.

$$rmsr = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (3.22)$$

Das Kriterium *rmsr* läßt sich mit Hilfe des Mittelwertes der Beobachtungen normieren, mit dem normierten Wert $rmsr_{norm}$ kann eine Einschätzung der Größenordnung der Schätzabweichungen durchgeführt werden (Loague & Green, 1991):

$$rmsr_{norm} = \frac{1}{\bar{y}} \sqrt{\frac{1}{N-1} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad \text{falls } \bar{y} \neq 0. \quad (3.23)$$

3.2.3.3 Robuste Lageschätzung der Schätzabweichungen

Eine robuste Lageschätzung der Schätzabweichungen (robust estimation of residuals; *rer*) kann ebenfalls aufschlussreich für die Beurteilung von Schätzverfahren sein, da extreme Schätzabweichungen dabei stärker berücksichtigt werden als bei *me_r*, den Schätzwert jedoch nicht beliebig dominieren können wie bei *rmsr*. Die robuste Schätzung der Schätzabweichungen berechnet sich als:

$$rer = T_N((y_1 - \hat{y}_1), \dots, (y_N - \hat{y}_N)) \quad (3.24)$$

Es ist zu erwarten, dass *rer* schärfer ist als *robust bias*, da ein *rer* Wert von Null hinreichend ist für einen Wert von *robust bias* gleich Null, aber nicht umgekehrt.

3.2.3.4 Regressionsparameter

Die Güte von Schätzungen kann auch mit verschiedenen Parametern der Regressionsgerade zwischen den geschätzten und den gemessenen Werten charakterisiert werden:

Bestimmtheitsmaß

Das Ausmaß des linearen Zusammenhanges kann mit Hilfe des Bestimmtheitsmaßes der Regressionsgeraden (coefficient of determination of the regression line; *cdr*) charakterisiert werden (Janssen & Heuberger, 1995).

$$cdr = \frac{\left(\sum_{i=1}^N (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}}) \right)^2}{\sum_{i=1}^N (\hat{y}_i - \bar{\hat{y}})^2} \quad \text{mit } 0 \leq cdr \leq 1 \quad (3.25)$$

Ein Bestimmtheitsmaß nahe eins besagt, dass die Schätzwerte gut mit einer linearen Regressionsgerade mit einer Steigung ungleich Null wiedergegeben werden können, d. h. dass ein *enger linearer Zusammenhang* zwischen den geschätzten und den gemessenen Werten besteht. Dies bedeutet jedoch nicht, dass die Schätzung selbst schon optimal ist. Eine Regressionsgerade mit einem Bestimmtheitsmaß von eins kann einen Achsenabschnitt ungleich Null und eine Steigung ungleich eins aufweisen. Daher müssen außer dem Bestimmtheitsmaß *cdr* auch die Regressionskoeffizienten zur Beurteilung eines Schätzverfahrens herangezogen werden.

Steigung und Achsenabschnitt

Der Achsenabschnitt und die Steigung der Regressionsgeraden charakterisieren bei einer Abweichung von den Werten Null für den Achsenabschnitt und eins für die Steigung den Bias (siehe Abb. 3.3b) bzw. den systematischen Fehler (siehe Abb. 3.3e) einer Schätzung. Diese Fehlertypen stellen wichtige Eigenschaften des zu beurteilenden Schätzverfahrens dar und geben Aufschluss über etwaige Mängel des verwendeten Modells, z. B. wenn die Zusammenhänge zwischen zu schätzender und den erklärenden Variablen im Modell nicht ausreichend beschrieben sind. Zur Bestimmung der Koeffizienten der Regressionsgeraden wird beispielhaft ein robustes Verfahren von Tukey vorgestellt. Die Methode besteht darin, die Steigung durch den Median der Steigungen von Datenpaaren zu berechnen (Tukey, 1977)

$$b = \text{med}_{1 \leq i \leq \frac{N+1}{2}} \left\{ \frac{\hat{y}_{N+1-i} - \hat{y}_i}{y_{N+1-i} - y_i} \right\}, \quad y_1 \leq y_2 \leq \dots \leq y_N \quad (3.26)$$

Der Index i läuft hierbei nur bis zur größten ganzen Zahl kleiner oder gleich $\frac{N}{2}$, da andernfalls redundante Datenpaare gebildet werden. Weiterhin ist der Nenner $y_{N+1-i} - y_i$ gleich Null, falls N ungerade ist und $i = \frac{(N+1)}{2}$.

Der Achsenabschnitt berechnet sich mittels der Steigung b als:

$$a = \text{med}_{1 \leq i \leq N} (\hat{y}_i - by_i) \quad (3.27)$$

Diese Methode hat den Vorteil, dass der Bruchpunkt $1/6$ beträgt (Nagel, 1994), d. h. dass bis zu $1/6$ Ausreißer im Datensatz enthalten sein können, ohne dass sich Steigung und Achsenabschnitt beliebig verändern können.

3.2.4 Referenzverfahren

Unter einem Referenzverfahren wird ein sehr einfaches Schätzverfahren verstanden, z. B. die Schätzung aller Werte mittels einer einzigen Konstanten. Bislang wurden Gütemaße beschrieben, die keinen Vergleich mit einem Referenzverfahren beinhalten. Derartige Gütemaße können zum relativen Vergleich von verschiedenen Schätzverfahren herangezogen werden. Es ist jedoch sinnvoll, die Güte eines Schätzverfahrens mit dem eines einfachen Referenzverfahrens zu vergleichen, um absolute Aussagen über die Güte des Verfahrens hinsichtlich der Schätzabweichung zu erhalten. Die Güte des Referenzverfahrens kann dann als Minimalanforderung an die zu beurteilenden Schätzverfahren verstanden werden. Durch die Einbeziehung des Referenzverfahrens ist es möglich, eine Intervallskala (Stevens, 1946) für die jeweiligen Gütemaße zu definieren, welche dann für die Bewertung der Schätzverfahren bereitsteht.

Allgemein können Gütemaße, die auf Koeffizienten der Schätzabweichungen (coefficients of residual errors; *core*) aufbauen, wie folgt beschrieben werden:

$$\text{core} = 1 - \frac{f(y_1 - \hat{y}_1, \dots, y_N - \hat{y}_N)}{f(y_1 - a, \dots, y_N - a)} \quad \text{mit } y_{\min} < a < y_{\max} \quad (3.28)$$

wobei f eine geeignete Funktion darstellt, etwa $f(\xi_1, \dots, \xi_N) = \sum_{i=1}^N \xi_i^2$, $f(\xi_1, \dots, \xi_N) = \max_i (\xi_i)$.

Der Parameter a wird dabei z. B. als Lösung folgender Minimax-Bedingung ermittelt:

$$\max_i (y_i - a) = \min_i (y_i - a) \quad (3.29)$$

Die obere Schranke von *core* ist immer 1. Negative Werte treten dann auf, wenn die Schätzungen \hat{y}_i schlechter sind als die des Referenzverfahrens. Die Schätzung ist umso besser, je näher *core* an eins ist. Es gibt verschiedene Gütefunktionen, die jeweils zu unterschiedlichen Maßen und Charakterisierungen führen:

Der Koeffizient der maximalen Schätzabweichung (coefficient of maximum residual error; *comare*) sei definiert als:

$$comare = 1 - \frac{\max\{|y_1 - \hat{y}_1|, \dots, |y_N - \hat{y}_N|\}}{0.5 |\max\{y_1, \dots, y_N\} - \min\{y_1, \dots, y_N\}|} \quad (3.30)$$

Mit diesem Gütemaß kann überprüft werden, ob die *größte* Schätzabweichung des zu überprüfenden Schätzverfahrens größer ist als die größte Schätzabweichung einer geeigneten konstanten Schätzung. Der Koeffizient der *zentralen* Schätzabweichung (coefficient of the central residual error; *cocare*; EF* nach Zacharias et al., 1996)

$$cocare = 1 - \frac{\text{med} \{|y_1 - \hat{y}_1|, \dots, |y_N - \hat{y}_N|\}}{\text{med} \{|y_1 - \text{med}\{y_1, \dots, y_N\}|, \dots, |y_N - \text{med}\{y_1, \dots, y_N\}|\}} \quad (3.31)$$

vergleicht den Median der Schätzabweichung eines Schätzverfahrens mit der median-deviation (cf. Sachs, 1997) der Schätzabweichung.

Die Distanzfunktionen von *cocare*, Median und Median der Medianabweichung, weisen einen Bruchpunkt von $\epsilon^* = 0.5$ auf (Huber, 1981), daher ist *cocare* hinsichtlich Ausreißern ein sehr stabiles Kriterium, aber empfindlich gegenüber kleinen Änderungen der Werte.

Der Koeffizient der mittleren quadrierten Schätzabweichung (coefficient of mean square residual error; *comsre*) berechnet sich als (Nash & Sutcliffe, 1970; Agterberg, 1984; Loague & Green, 1991)¹:

$$comsre = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (3.32)$$

Das Kriterium *comsre* zielt auf die mittlere Schätzabweichung des zu beurteilenden Schätzverfahrens und vergleicht diese mit der mittleren Schätzabweichung eines Referenzschätzers. Der Referenzschätzer ist in diesem Fall der Mittelwert. Die mittlere Schätzabweichung des Referenzschätzers entspricht daher der Varianz der Messungen. Sie ist jedoch aufgrund des minimalen Bruchpunktes von Null (Hampel, 1986) sehr anfällig gegenüber Ausreißern, so dass diese das Gütemaß stark beeinflussen.

Der Koeffizient der robusten Schätzung der Schätzabweichungen (coefficient of robust estimation of residual error; *corore*) berechnet sich als

$$corore = 1 - \frac{T_N(|y_1 - \hat{y}_1|, \dots, |y_N - \hat{y}_N|)}{T_N(|y_1 - T_N(y_1, \dots, y_N)|, \dots, |y_N - T_N(y_1, \dots, y_N)|)} \quad (3.33)$$

Das robuste Gütekriterium *corore* weist gegenüber *comsre* sowohl eine geringere Ausreißerempfindlichkeit (*gross error sensitivity*) als auch eine geringere Rundungsempfindlichkeit (*local shift sensitivity*) auf, Einzelwerte können das Gütemaß weniger beeinflussen. Die Schätzung der Güte ist daher sicherer. Gegenüber dem Gütekriterium *comsre* hat *corore* den Vorteil einer geringeren Rundungsempfindlichkeit der Koeffizienten.

¹ Diese Autoren nennen das Kriterium 'modelling efficiency' ME

3.3 Kreuzvalidierung

Die vorliegende Arbeit beschränkt sich auf räumliche Schätzungen, die auf Messwerten an einzelnen Punkten im Untersuchungsgebiet und flächenhaft vorgehaltenen Daten, z. B. aus Kartenwerken, basieren. Eine Güteabschätzung dieser räumlichen Schätzungen kann deshalb nur auf Stichproben der Grundgesamtheit des Untersuchungsgebietes aufbauen, soll aber eine begründete Aussage über die Güte der Schätzung im gesamten Untersuchungsgebiet liefern. Gesucht ist deshalb eine Methode, die eine Güteabschätzung im gesamten Gebiet auf der Grundlage der Messdaten ermöglicht, d. h. die die Frage beantwortet: ‘Wie gut ist die Schätzung im Gebiet?’

Eine Möglichkeit die Güte zu schätzen, besteht darin, den Datensatz und die jeweiligen Schätzverfahren einer Kreuzvalidierung (engl. *cross validation*, *predictive sample reuse*) zu unterziehen. Die Kreuzvalidierung ermöglicht den Vergleich zwischen Messwert und Schätzwert an allen Punkten im Raum, an dem Messungen vorgenommen wurden. Dabei werden alle Schätzwerte unabhängig vom jeweiligen Messwert gewonnen. Das Verfahren kann daher auch als Simulation einer neu zum Datensatz hinzugenommenen Messung aufgefasst werden.

Die Methode der Kreuzvalidierung ist schon lange bekannt, erlebt aber neuerdings durch die leichte Verfügbarkeit großer Rechenleistungen durch die moderne Computertechnologie eine Renaissance. Die Originalidee besteht darin, die vorhandenen Daten in zwei etwa gleiche Teile aufzuteilen und mit der einen Hälfte der Daten (Trainingsmenge) ein Modell zunächst optimal an die Daten anzupassen, um dann mit diesem Modell die andere Hälfte (Testmenge) vorherzusagen. Gewöhnlich ist dabei die Anpassung bei der Trainingsmenge besser als bei der Testmenge, da das Modell ja optimal an die Trainingsmenge angepasst wird. Sachlich besteht keine Notwendigkeit, die Datenmenge in zwei gleiche Mengen aufzuteilen. Aufgrund geringerer Rechenkosten ist deshalb folgende Methode in der Praxis weit verbreitet (Stone, 1975, Efron, 1983):

1. Aufteilung des Datensatzes in eine Trainingsmenge

$$\mathbf{X}_{(i)} = (\mathbf{x}_1, \dots, \mathbf{x}_{i-1}, \mathbf{x}_{i+1}, \dots, \mathbf{x}_n) \quad (3.34)$$

und Testmenge (\mathbf{x}_i), wobei die Testmenge nur aus einem Element, nämlich dem Vektor der i -ten Beobachtung

$$(\mathbf{x}_i) = (\mathbf{u}_i, y_i) \text{ besteht}$$

2. Modellbildung und -optimierung mit der Trainingsmenge

$$\mathbf{y}_{(i)} = \eta_{\mathbf{x}_{(i)}}(\mathbf{u}_{(i)}) + \boldsymbol{\varepsilon}_{(i)} \quad (3.35)$$

mit

$$\boldsymbol{\varepsilon}_{(i)}^T \boldsymbol{\varepsilon}_{(i)} = \min \quad (3.36)$$

3. Anwendung des optimierten Modells auf die Testmenge und Berechnung des Schätzfehlers

$$\hat{y}_i = \eta_{\mathbf{x}_{(i)}}(\mathbf{u}_i) \quad (3.37)$$

$$y_i - \hat{y}_i = y_i - \eta_{\mathbf{x}(i)}(\mathbf{u}_i) \quad (3.38)$$

4. Wiederholung der Schritte 1-3 bis jedes Element unabhängig vorhergesagt wurde

5. Mit Gl. 3.38 können diverse Schätzfehlerstatistiken z. B. nach Gl. 3.39 berechnet werden.

3.3.1 Schätzung der Güte im Untersuchungsgebiet mit der Kreuzvalidierung

Die Schätzfehler, die mit Hilfe der Kreuzvalidierung ermittelt werden, werden im Folgenden ‘kreuzvalidierte Schätzfehler’ genannt und können mit den in Kap. 3.1 vorgestellten Gütekriterien charakterisiert werden. Der mittlere kreuzvalidierte Schätzfehler ist der Erwartungswert der durch die Kreuzvalidierung ermittelten Schätzfehler (Efron, 1983, 1993):

$$err(\mathbf{X}, F) \equiv \hat{err}^{cross} = E_{0F} \left\{ Q \left[Y_0, \eta_{\mathbf{x}}(U_0) \right] \right\} = \frac{1}{n} \sum_{i=1}^n \left(y_i - \eta_{\mathbf{x}(i)}(\mathbf{u}_i) \right) \quad (3.39)$$

Je kleiner die kreuzvalidierten Schätzfehler, desto eher besitzt das Modell Gültigkeit für die Testelemente der Stichprobe und desto vollständiger werden die Systemvariablen, welche die Stichprobe beschreiben, erfasst (Sachs, 1997). Die kreuzvalidierten Schätzfehler können deshalb zur Charakterisierung der Validität der Stichprobe verwendet werden. Unter der Annahme, es liege eine repräsentative Stichprobe vor, stellen die kreuzvalidierten Schätzfehler der Stichprobe eine Schätzung für alle Punkte im Untersuchungsgebiet dar, die nicht beprobt wurden. Daher sind die kreuzvalidierten Schätzfehler eine Grundlage für die Schätzung der Güte des Schätzverfahrens im Untersuchungsgebiet. Der Unterschied zwischen Kreuzvalidierung, *bootstrap* und *jackknife* wird in Anhang 2 erläutert.

3.4 CART - Classification and Regression Trees

3.4.1 Einleitung und Übersicht

Das CART-Verfahren ist ein Algorithmus zur Beschreibung des Zusammenhangs zwischen Eingangsgrößen und einer Zielgröße (Breiman *et al.*, 1984). Im vorliegenden Falle stellen die Variablen Nitratgehalt im Boden und die Korngrößenfraktionsanteile des Bodens eine Zielgröße dar, Bodenschätzungsparameter, Anbaufrucht und geologischer Untergrund sind als Eingangsgrößen anzusehen.

Die Eingangsgrößen und die Zielgröße können beim CART-Verfahren unterschiedlichen Skalenarten (Stevens, 1946) angehören. Bei einer nominalskalierten¹ Zielgröße, z. B. Bodentyp, spricht man von einem Klassifikationsbaum (*classification tree*), bei einer reellen Zielgröße, z. B. Nitratgehalt, von einem Regressionsbaum (*regression tree*). Beim CART-Verfahren werden wie bei der Regression zwei Ziele verfolgt: bestmögliche Vorhersage der Zielgröße für zukünftige Messungen und Verständnis der strukturellen Beziehungen zwischen den Eingangsgrößen und der Zielgröße.

Der Begriff CART wird synonym sowohl für den Algorithmus als auch für das entsprechende Softwareprogramm von Salford Systems, San Diego, verwendet. Vom Datensatz ausgehend, wird zunächst ein Entscheidungsbaum gebildet, der dann zur Klassifikation neuer Daten herangezogen wird. Der Algorithmus selbst beruht auf dem sogenannten Prinzip der *binären rekursiven Partitionierung*, d. h.,

¹ Bei nominalskalierten Größen existiert keine objektive Reihenfolge der Merkmalsausprägung (Stevens, 1946)

der Entscheidungsbaum wird erzeugt, indem die Stichprobe L zunächst in zwei Teilmengen geteilt wird und diese Teilmengen ihrerseits fortlaufend wieder in jeweils zwei Teilmengen aufgeteilt werden. Das Verfahren wird als rekursiv bezeichnet, da jede Menge, die aus einem Teilungsprozess entstanden ist, wiederum Ausgangsmenge einer neuen Teilung ist.

Die Aufteilung wird höchstens so lange durchgeführt, bis keine weitere Teilung mehr sinnvoll ist d. h. alle Elemente einer Teilmenge die gleiche Merkmalsausprägung der Zielgrößen, wie z. B. den gleichen Bodentyp oder den gleichen Stickstoffgehalt, besitzen. Ein zusätzliches Abbruchkriterium ist die Unterschreitung einer Mindestanzahl von Elementen der Teilmengen. In Anlehnung an die Graphentheorie (Bodendiek & Henn, 1990) wird jede Teilmenge als *Knoten* bezeichnet; die Stichprobe L , aus der der Entscheidungsbaum erzeugt wird, heißt *Wurzelknoten* t_1 , eine weiter nicht mehr unterteilte Menge *Endknoten*. Jede Aufteilung wird nach ihrer Fähigkeit beurteilt, die im Knoten enthaltenen Klassen möglichst gut zu trennen. Ein Knoten heißt *rein*, wenn er nur noch Objekte mit gleicher Merkmalsausprägung d. h. gleichem Wert der Zielgröße enthält. Die Regel nach der ein Knoten aufgrund einer Eingangsgröße x geteilt wird, heißt Teilungsregel oder *Split*. Mathematisch wird die Teilungsregel durch eine Indikator-Funktion $I_C(x)$ beschrieben. Alle Datenelemente, bei denen $I_C(x)=1$ ist, werden in eine sogenannten *linke* Teilmenge t_L eingegliedert, alle anderen in die rechte, t_R . Die Menge C der Indikator-Funktion $I_C(x)$ ergibt sich wie folgt:

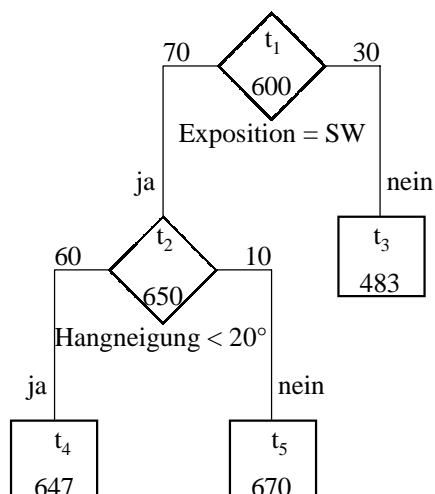
$$x \text{ nominalskaliert: } C = \{c_1, \dots, c_n\} \quad \text{d. h. } I_C(x)=1 \Leftrightarrow x \in C; \quad I_C(x)=0 \Leftrightarrow x \notin C \quad (3.40)$$

$$x \text{ reell: } C = \{x \mid x \leq c\} \quad \text{d. h. } I_C(x)=1 \Leftrightarrow x \leq c; \quad I_C(x)=0 \Leftrightarrow x > c \quad (3.41)$$

Abb. 3.6 zeigt das Schema eines Regressionsbaumes mit drei Knotenebenen.

Abb. 3.5: Schema eines Entscheidungsbaumes

Dargestellt ist ein Regressionsbaum zur Schätzung des Niederschlages aus den Variablen Exposition und Hangneigung eines Messortes (hypothetischer Datensatz). Quadrate kennzeichnen Splits, Rauten Endknoten. Die Zahlen beiderseits der Rauten bezeichnen die Anzahl der Elemente auf dem jeweiligen Ast. Die Zahlen in den Rauten und Quadraten stehen für den Mittelwert aller Elemente in dem jeweiligen Knoten. Ausgangslage ist eine Stichprobe L mit einem Umfang von 100 Elementen mit einem mittleren Niederschlagswert von 600 mm. Die Elemente der Stichprobe L (=Wurzelknoten t_1) werden durch die Aufteilungsregel: „Ist die Exposition südwestexponiert?“ in zwei Teilmengen t_2 und t_3 geteilt. Teilmenge t_2 hat 70 Elemente mit einem Mittelwert von 650 mm, Teilmenge t_3 hat einen Umfang von 30 mit einem mittleren Niederschlagswert von 483. Teilmenge t_2 wird ihrerseits nochmals mit der Aufteilungsregel: „Ist die Hangneigung kleiner 20 Grad?“ in die Teilmengen t_4 und t_5 untergliedert. 60 Elemente mit einem mittleren Niederschlag von 647 weisen eine kleinere Hangneigung als 20 Grad auf; 10 Untersuchungsstellen mit einem Mittelwert von 670 mm haben eine größere Hangneigung. Der Wurzelknoten t_1 wurde insgesamt durch zwei Aufteilungen in die Endknoten t_3 , t_4 und t_5 aufgeteilt.



Der auf diese Art erzeugte Baum wird als Maximalbaum bezeichnet. Maximalbäume sind jedoch nicht zur Klassifikation oder Schätzung geeignet, da sie optimal an die Lernstichprobe angepasst sind und bei Anwendung auf einen anderen Datensatz i. A. sehr schlechte Ergebnisse liefern.

Es hat sich als effektiv herausgestellt, zuerst einen Maximalbaum zu generieren und diesen danach sukzessive zu *stutzen* (engl.: Pruning)², so dass eine Folge von geschachtelten Teilbäumen entsteht. Aus dieser Folge wird der Baum mit optimaler Größe ermittelt. Bei der Ermittlung der optimalen Baumgröße gehen die Zahl der Endknoten und der geschätzte Vorhersagefehler ein. Für die Schätzung des Vorhersagefehlers wird eine modifizierte Kreuzvalidierung verwendet. In den folgenden Abschnitten werden die oben kurz skizzierten Grundprinzipien, die für Klassifikations- und Regressionsbäume gleichermaßen gelten, am Beispiel der später verwendeten Regressionsbäume mit der kleinsten Quadrate Anpassung detaillierter dargestellt. Dabei wird die gleiche Nomenklatur angewendet, die in Kap. 1.5 eingeführt wurde.

3.4.2 Kleinste Quadrate Regression

Im Regressionsfall besteht ein Datenelement aus einem Vektor $\mathbf{x} = (\mathbf{u}, y)$, wobei \mathbf{u} der Vektor der Eingangsvariablen und y eine davon abhängige, reelle Variable darstellt. Die Vorhersagefunktion $\eta(\mathbf{u})$ ist eine reelle Funktion auf dem Zahlenraum der Eingangsvariablen. Regression ist die Ermittlung einer Funktion $\eta_L(\mathbf{u})$ auf der Basis einer Stichprobe $L = \{(\mathbf{u}_1, y_1), \dots, (\mathbf{u}_N, y_N)\}$ mit N Datenelementen.

Die Funktionswerte der Regressionsfunktion $\eta_L(\mathbf{u})$ der Elemente eines Knotens t berechnen sich durch den Mittelwert der Zielgröße aller Elemente des Knotens (Breiman *et al.*, 1984):

$$\eta_L(t) = \bar{y}(t) = \frac{1}{N(t)} \sum_{y_n \in t} y_n \quad (3.42)$$

Die Güte der Regressionsfunktion $\eta(\mathbf{u})$ wird mittels eines Fehlermaßes beurteilt, idealerweise des Erwartungswertes $R^*(\eta)$ der quadrierten Differenzen zwischen Beobachtung und Schätzwert (Breiman *et al.*, 1984):

$$R^*(\eta) = E(Y - \eta(\mathbf{X}))^2 \quad (3.43)$$

Man beachte, dass R^* von der Wahl der Skala abhängig ist, mit der die Zielgröße gemessen wurde (Breiman *et al.*, 1984). Es gibt mehrere Möglichkeiten, um den Fehler R^* des Schätzverfahrens $\eta = \eta_L(\mathbf{u})$ zu schätzen:

Ein üblicher, aber relativ schlechter Schätzer von R^* ist der *Resubstitutionsfehler* R_L der Stichprobe L (Breiman *et al.*, 1984). Zu seiner Berechnung wird der Datensatz sowohl für die Ableitung des Schätzverfahrens, als auch für die Gütebestimmung verwendet (Breiman *et al.*, 1984):

$$R_L = \frac{1}{N} \sum_{n=1}^N (y_n - \eta_L(\mathbf{u}_n))^2 \quad (3.44)$$

Von den vielen Möglichkeiten der Schätzung von R^* sollen außer dem Resubstitutionsfehler zwei weitere Schätzer eingeführt werden. Zur Ermittlung des *Teststichprobenfehlers* $R_{L,ts}$ (Breiman *et al.*,

² Der von Breiman *et al.* ursprünglich begangene Weg, Stoppregeln beim Baumwachstum einzuführen, führte in eine methodische Sackgasse.

1984) wird der zur Verfügung stehende Datensatz L per Zufall in eine Lernstichprobe $L^{(i)}$ und eine Teststichprobe L_1 mit $L^{(i)} \cap L_1 = \emptyset$ und $L^{(i)} \cup L_1 = L$ geteilt. Die Güteabschätzung wird unabhängig von der Lernstichprobe mit der Testmenge L_1 durchgeführt (Breiman *et al.*, 1984):

$$R_{L_1, ts} = \frac{1}{|L_1|} \sum_{(\mathbf{u}_n, y_n) \in L_1} (y_n - \eta_{L^{(i)}}(\mathbf{u}_n))^2 \quad (3.45)$$

Die Teststichproben-Schätzung $R_{L_1, ts}$ ist die ursprüngliche Idee der Kreuzvalidierung wie sie von Stone (1974), Geisser (1975) und Efron (1983) beschrieben wurde (Kap. 3.3).

Bei der Schätzung durch den von Breiman (1984) beschriebenen v -fachen *Kreuzvalidierungsfehler* $R_{L_i, cv}$ wird der Datensatz L in v zufällige Teilmengen L_1, \dots, L_v mit möglichst gleichem Umfang und folgenden Eigenschaften aufgeteilt: $L_i \cap L_k = \emptyset$, für alle $i, k \in \{1, \dots, v\}, i \neq k$ und $L_1 \cup L_2 \cup \dots \cup L_v = L$. Für jedes $i = 1, \dots, v$ wird auf Basis der Lernstichprobe $L^{(i)} = L - L_i$ ein Regressionsbaum $\eta_{L^{(i)}}$ gebildet und die v -fache Kreuzvalidierungsfehler $R_{L_i, cv}$ ermittelt (Breiman *et al.*, 1984):

$$R_{L_i; cv} = R_{L_1, \dots, L_v; cv} = \frac{1}{N} \sum_{i=1}^v \sum_{(\mathbf{u}_n, y_n) \in L_i} (y_n - \eta_{L^{(i)}}(\mathbf{u}_n))^2 \quad (3.46)$$

Dieses Verfahren ist eine Verallgemeinerung der in Anhang 2 beschriebenen Kreuzvalidierung, da das von Breiman angewendete Verfahren bei einelementigen Teststichproben L_v , d. h. $v = N$, mit dem in Kap. 3.3 ausgeführten identisch ist.

Die Schätzung von $R^*(\eta)$ hängt ebenso wie $R^*(\eta)$ selbst von der Wahl der Skala für Y ab; sie kann skalenunabhängig gemacht werden, indem $R^*(\eta)$ mit Hilfe der Fehlerschätzung für den Erwartungswert $\mu = \mu(u) = E(Y)$ der Zielvariable normiert wird; die entsprechende Schätzung heißt *relativer Fehler* (Breiman *et al.*, 1984):

$$RE^*(\eta) = \frac{R^*(\eta)}{R^*(\mu)} \text{ mit } R^*(\mu) = E(Y - \mu)^2 = E(Y - E(Y))^2 \quad (3.47)$$

3.4.3 Teilung eines Knotens

Die Teilung eines Knotens beim Bildungsprozess der Regressionsbäume wird nach folgender Regel durchgeführt: Die beste Teilung s^* eines Knotens t ist diejenige aus der Menge aller möglichen Teilungen S , die den Resubstitutionsfehler $R_L(t)$ des Knotens t minimiert, so dass für die durch die Teilung entstehenden zwei Teilmengen t_L und t_R gilt (Breiman *et al.*, 1984):

$$\Delta R(s^*, t) = \max_{s \in S} \Delta R(s, t) \text{ mit } \Delta R(s, t) = R(t) - R(t_L) - R(t_R) \quad (3.48)$$

d. h. die beste Teilung eines Knotens t ist die, welche den Knoten t anhand der Merkmalsausprägungen \mathbf{x} am besten in eine Gruppe mit hohen und eine Gruppe mit niedrigen Werten der Zielgröße aufteilt.

Die Teilung wird nach den oben beschriebenen Regeln i. A. so lange fortgesetzt, bis für jeden Endknoten $t \in T_{\max}$ die Anzahl der Elemente $N(t)$ geringer ist als ein vorgegebener Wert N_{\min} . Gewöhnlich wird $N_{\min} = 5$ gewählt. Das nur aus theoretischer Sicht wichtige Abbruchkriterium bei Regressionsbäumen, dass nämlich alle Werte der Zielvariablen in einem Knoten den gleichen Wert haben, wird in der Praxis nur selten erfüllt.

3.4.4 Stutzen des Maximalbaumes (Pruning)

Der auf obige Weise entstandene Maximalbaum weist normalerweise eine wesentlich höhere Komplexität auf, als der Stichprobe angemessen ist. Da die Zahl der erfassten Variablen nicht ausreicht, um die Zusammenhänge der Datenstruktur vollständig zu beschreiben, ergeben sich zufällige Fehler, die sich in inhomogenen Endknoten manifestieren. Der CART-Algorithmus versucht mit den vorhandenen Variablen möglichst homogene Gruppen zu bilden, daher werden auch unsinnige Aufteilungsregeln bei der Bildung des Maximalbaumes verwendet. Aus diesem Grunde ist es erforderlich, die Komplexität zu reduzieren und nicht sinnvolle Splits zu verwerfen. Hierzu wird der Baum gestutzt. Die der Lernstichprobe angepasste Komplexität wird durch Kreuzvalidierung ermittelt.

Beim Pruning-Algorithmus des CART-Verfahrens wird durch sukzessives Stutzen des Maximalbaumes T_{\max} unter Verwendung des Fehler-Komplexitätsmaßes $R_{\alpha}(T)$ eine Folge von Teilbäumen $T_1 \succ T_2 \succ \dots \succ t_1$ erstellt. Das Zeichen \succ bedeutet, dass jeder Baum T_k mit $k > n$ ein Teilbaum des Baumes T_n darstellt; t_1 bezeichnet den Baum, der nur aus dem sogenannten Wurzelknoten d. h. der Stichprobe L besteht.

Das Fehler-Komplexitätsmaß $R_{\alpha}(T)$ eines Baumes T berechnet sich als Summe aus Resubstitutionsfehler des Baumes (Gl. 3.44) und eines Summanden, in den die Komplexität des Baumes eingeht (Breiman *et al.*, 1984):

$$R_{\alpha}(T) = R_L(T) + \alpha |T|, \quad 0 = \alpha_1 < \alpha_2 < \dots ; \quad (3.49)$$

mit $|T|$ = Anzahl der Endknoten des Baumes T

Der Faktor α heißt Komplexitätsparameter, weil mit steigenden α -Werten der Einfluss der Baumgröße $|T|$ auf $R_{\alpha}(T)$ stetig zunimmt. Die Folge von Teilbäumen $T_1 \succ T_2 \succ \dots \succ t_1$ entsteht, indem der Komplexitätsparameter α von Null beginnend stetig erhöht wird und für jedes α derjenige Teilbaum gesucht wird, der das Fehler-Komplexitätsmaß $R_{\alpha}(T)$ minimiert. Es ergeben sich Intervalle $[\alpha_k, \alpha_{k+1})$, so dass für alle α mit $\alpha_k \leq \alpha < \alpha_{k+1}$ der Baum T_k derjenige Teilbaum von T_{k-1} ist, der $R_{\alpha}(T)$ minimiert (Breiman *et al.*, 1984).

Aus der Sequenz der Teilbäume $T_1 \succ T_2 \succ \dots \succ t_1$ wird die optimale Baumgröße mit Hilfe der Kreuzvalidierung ermittelt. Dabei wird der Datensatz L in v zufällige, möglichst gleich große Teildatensätze L_v geteilt und die jeweils komplementären Mengen zum Gesamtdatensatz gebildet:

$$L^{(v)} = L - L_v \quad (3.50)$$

Als Wert von v wird von Breiman *et al.* (1984) mindestens 10 empfohlen, bei geringerer Zahl ist die Schätzung der Resubstitutionskosten deutlich ungenauer, bei höherer Zahl steigt der Rechenaufwand

bei nur unwesentlichen Verbesserungen der Schätzung. Für jede Menge $L^{(i)}$, $i = 1 \dots v$, wird auf oben geschilderte Weise unter Verwendung des Komplexitätsparameters α' zunächst ein Maximalbaum $T_{\alpha'_1}^{(i)}$ erstellt und dieser anschließend gestutzt, so dass für jedes $L^{(i)}$ eine Folge von Bäumen $T_{\alpha'_1}^{(i)}, T_{\alpha'_2}^{(i)}, \dots$ entsteht. Beim Stutzen werden jedoch für alle Bäume $T_{\alpha'_k}^{(1)}, \dots, T_{\alpha'_k}^{(v)}$ der gleiche Komplexitätsparameter α'_k verwendet. Diese berechnen sich nach folgender Formel von den Komplexitätsparametern α_k der Folge $T_{\alpha_1}, T_{\alpha_2}, \dots$, die auf Basis der Gesamtstichprobe L entwickelt wurde (Breiman *et al.*, 1984):

$$\alpha'_k = \sqrt{\alpha_k \cdot \alpha_{k+1}} \quad (3.51)$$

Auf Grund der gemeinsamen Komplexitätsparameter α'_k wird die Baumgröße bei der Berechnung des Fehler-Komplexitätsmaßes bei allen Teilbäumen $T_{\alpha'_k}^{(1)}, \dots, T_{\alpha'_k}^{(v)}$ gleich gewichtet. Für jedes k wird der v -fache Kreuzvalidierungsfehler $R_{L_i, cv}$ des Baumes T_{α_k} berechnet (Breiman *et al.*, 1984):

$$R_{L_i, cv}(T_{\alpha_k}) = \frac{1}{N} \sum_{p=1}^v \sum_{i \in L_p} (y_i - \eta_{T_{\alpha_k}, L^{(p)}}(i))^2 \quad (3.52)$$

Der Standardfehler des Schätzfehlers $R_{L_i, cv}(T_{\alpha_k})$ gibt Aufschluss über die Streuung der Schätzungen für $R^*(\eta)$. Er ist beim Baumschnitt bei der Auswahl des optimalen Baumes aus der Sequenz aller Teilbäume $T_{\alpha_1}, T_{\alpha_2}, \dots$ insofern von Bedeutung, als dieser üblicherweise mit Hilfe der *Standardfehler-Regel* bestimmt wird. Der Standardfehler $SE(R_{L_i, cv}(T_{\alpha_k}))$ berechnet sich wie folgt (Breiman *et al.*, 1984):

$$SE(R_{L_i, cv}(T_{\alpha_k})) = \sqrt{s^2 / N} \quad \text{mit} \quad (3.53)$$

$$s^2 = \frac{1}{N} \sum (y_n - \eta_{L^{(i)}}(\mathbf{u}_n))^2 - R_{L_i, cv}(T_{\alpha_k}) \quad (3.54)$$

Der optimale Baum ist der kleinste Baum T_{α_k} (d. h. der Baum mit kleinsten Zahl von Endknoten), der folgende Bedingung erfüllt (Breiman *et al.*, 1984):

$$R_{L_i, cv}(T_{\alpha_k}) \leq R_{L_i, cv}(T_{\alpha_0}) + SE \quad (3.55)$$

$$\text{mit } R_{L_i, cv}(T_{\alpha_0}) = \min_k R_{L_i, cv}(T_{\alpha_k}) \quad (3.56)$$

Der nach Breiman *et al.* (1984) optimale Regressionsbaum ist daher zwar mit Hilfe der Kreuzvalidierung entstanden, diese wurde aber lediglich zur Bestimmung der optimalen Baumgröße der Sequenz von Teilbäumen $T_{\alpha_1}, T_{\alpha_2}, \dots$ verwendet, die auf Grundlage der Gesamtstichprobe L entwickelt wurde. Das bedeutet, dass ein Funktionswert \hat{y}_i der Regressionsfunktion nicht unabhängig vom Datenelement

(\mathbf{u}_i, y_i) geschätzt wird. Aus diesem Grund wurden für die Schätzung der Nitratgehalte im Boden (Fallbeispiel 1) echte Kreuzvalidierungsschätzungen durchgeführt, indem sukzessive ein Datenelement aus der Stichprobe ausgeblendet wurde, mit dem auf diese Art veränderten Datensatz der CART-Algorithmus durchgeführt wurde und anschließend der unabhängig geschätzte Wert des Datenelements ermittelt wurde. Da diese Vorgehensweise auch in der neuesten zur Verfügung stehenden CART Programmversion 4.0.0.20 nicht implementiert ist, musste diese echte Kreuzvalidierung manuell ausgeführt werden und konnte daher nur beim beschränkten Datensatz des Nitratgehaltes im Boden durchgeführt werden.

3.4.5 Eigenschaften der *classification and regression trees*

Von mehreren Autoren werden vorteilhafte Eigenschaften des *classification and regression tree*-Algorithmus von Breiman et. al. herausgestellt:

1. *Classification and regression trees* erfordern keine Vorauswahl der Vorhersagevariablen, sie können sogar explizit hierfür herangezogen werden (Kuhnert et al., 2000)
2. Entscheidungsbäume sind parameterfrei, Linearitäts- und Verteilungsannahmen brauchen nicht gemacht werden (Breiman et al., 1984). Daher sind *classification and regression trees* resistent gegenüber besonders kleinen oder großen Zahlenwerten (jedoch im engen Sinne nicht robust). Daher entfällt die Notwendigkeit zur Datentransformation, wie auch alle anderen Datenvorbehandlungsmethoden. Aus diesem Grund sind komplexe Datenstrukturen bei der Anwendung von CART unproblematisch, im Gegensatz zu parametrischen Verfahren, bei denen multivariate Verteilungsannahmen nur schwer verifiziert werden können
3. *Classification and regression trees* erlauben die Kombination von nominalskalierten (Stevens, 1946) und stetigen Daten (Breiman, et al., 1984)
4. Beliebige Datentransformationen sind möglich; Linearkombinationen der Variablen sind nicht zulässig (Steinberg u. Colla, 1997)
5. Entscheidungsbäume lassen sich einfach für die Klassifizierung neuer Objekte verwenden und leicht interpretieren (Fisher, 1990, Griffin et al. 1997, Emmons et al., 1999)
6. *Classification and regression trees* reagieren flexibel auf heterogene Merkmale und können daher den Einfluss von Ausreißern verringern (Steinberg u. Colla, 1997)
7. Datensätze mit *missing values* können zur Konstruktion von Entscheidungsbäumen verwendet werden (Steinberg u. Colla, 1997)

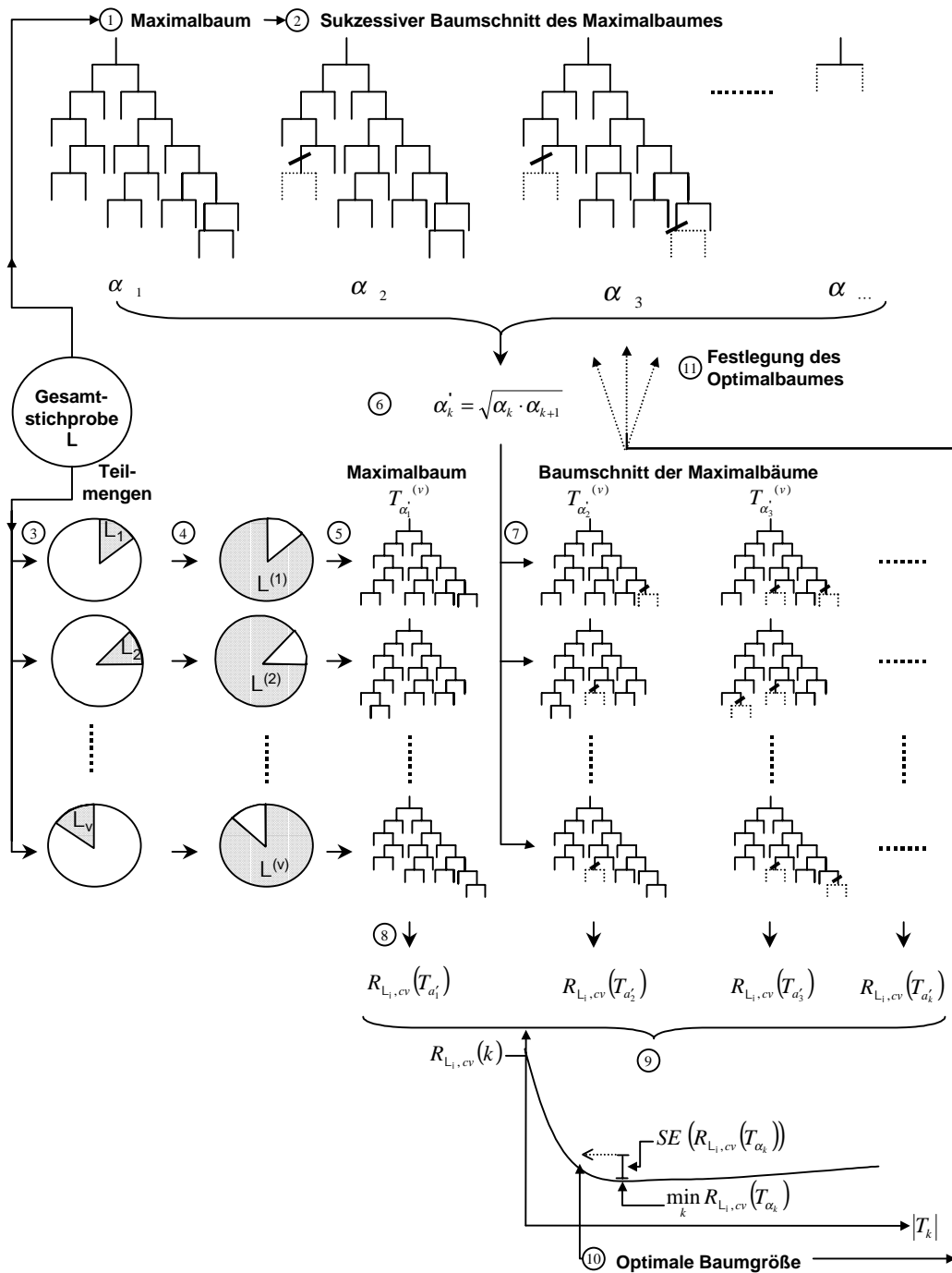


Abb. 3.6: Schematische Darstellung des CART Baumschnitts

Dargestellt sind die wesentlichen Schritte des Baumschnitts (Pruning). Von der gesamten Stichprobe wird im ersten Schritt ein Maximalbaum erstellt ① und dieser durch fortlaufende Erhöhung des Komplexitätsparameters α sukzessive beschnitten ② bis nach dem letzten Schritt nur noch der Wurzelknoten übrig bleibt. Danach wird die Stichprobe zufällig in v möglichst gleich große Teilmengen geteilt ③ und jeweils die Komplementmenge zur Gesamtstichprobe ermittelt ④. Die Komplementmengen dienen als Lernstichproben, um jeweils Maximalbäume zu generieren ⑤, die im Anschluss beschnitten werden ⑦. Dabei wird für alle v Maximalbäume ein gemeinsamer Komplexitätsparameter α' verwendet, der von den Komplexitätsparametern des Maximalbaumes abgeleitet wurde ⑥. Die kreuzvalidierten Schätzfehler $R_{L_i, cv}(T_{\alpha'_k})$ werden berechnet, indem der Schätzfehler über alle Teststichproben L_v gemittelt wird ⑧. Aus der Darstellung des Kreuzvalidierungsfehlers als Funktion der Baumgröße $|T_k|$ ⑨ werden die kleinstmöglichen Bäume $T_{\alpha'_j}^{(1) \dots (v)}$ ermittelt, deren Kreuzvalidierungsfehler kleiner sind als der minimale Kreuzvalidierungsfehler plus dessen Standardfehler ⑩. Der optimale Baum ist der Baum $T_{\alpha'_j}$ aus der Folge der Teilbäume auf der Grundlage der Gesamtstichprobe L .

3.5 Schätzung des Nitratgehaltes

3.5.1 Schätzung des Nitratgehaltes mit CART

Die Schätzung des Nitratgehaltes wurde in jeder der drei Tiefenstufen mit CART Version 4.0.0.20 von Salford Systems durchgeführt. Dabei wurden die in Kapitel 2.2 genannten Landnutzungsdaten, die Hangposition und der bewirtschaftende Betrieb als nominalskalierte erklärende Variable eingelesen und verarbeitet. Die Parametervoreinstellungen sind jeweils am Ende der einzelnen Ausgabeprotokolle aufgeführt. Die Ausgabeprotokolle sind auf der beigelegten CD-ROM enthalten („CART Ausgabeprotokoll Nitrat.doc“).

3.5.2 Vergleichsverfahren zur Schätzung der Nitratwerte im Boden

Die Ergebnisse der classification and regression trees werden mit Hilfe der in Kapitel 3.2 dargestellten Gütemaße mit den Ergebnissen einfacher Vergleichsverfahren verglichen um die Leistungsfähigkeit der *classification and regression trees* einschätzen zu können. Als einfache Vergleichsverfahren zur Schätzung der Nitratwerte wurden nachfolgend erläuterte Schätzverfahren angewendet. Die Berechnungen der Vergleichsverfahren (nächste Nachbarn Verfahren mit Varianten und Gruppenmittelwertverfahren) wurden mit Hilfe von MS Excel durchgeführt.

3.5.2.1 Nächster Nachbar

Mit Hilfe der Rechts- und Hochwerte der Beprobungsorte wurde für jeden Beprobungsort der nächste Nachbar ermittelt. Hierfür wurde das euklidische Distanzmaß verwendet. Als Kürzel des Verfahrens wird ‚NN‘ verwendet. Die Rechts- und Hochwerte der Probenahmestellen sind in Anhang 7 dokumentiert.

3.5.2.2 Nächster Nachbar mit gleicher Kulturart

Für die Bestimmung des nächsten Nachbarn gleicher Kulturart wurde zusätzlich zu den Ortskoordinaten die aktuelle Kulturart entsprechend der Datenerhebung der LUFA Augustenberg herangezogen. Am Beprobungstermin lagen acht verschiedene Kulturarten bzw. Bearbeitungszustände (Winterbrache, Stilllegung; siehe Anhang 7) vor. Das Verfahren wird nachfolgend mit ‚NN-Veg‘ bezeichnet.

3.5.2.3 Nächster Nachbar mit gleicher Hanglage

Bei der Variante nächster Nachbar gleicher Hanglage wurde die Hangposition des Beprobungsortes hinzu gezogen. Die Bestimmung erfolgte durch die LUFA Augustenberg. Dabei wurde seitens der LUFA zwischen den drei Hangpositionen Hang-Toplage, Hangmitte und Hangfuß unterschieden. Als Abkürzung wird ‚NN-Hang‘ verwendet. Die Hangpositionen sind ebenfalls in Anhang 7 vermerkt.

3.5.2.4 Schätzung mit Hilfe von Gruppenmittelwerten

Bei den Schätzungen mit Hilfe von Gruppenmittelwerten wurde die gesamte Stichprobe mit Hilfe der Merkmalsausprägungen einer erklärenden Variable in entsprechende Gruppen aufgeteilt und der Mittelwert dieser Gruppen als Schätzwert für alle Schläge mit gleicher Merkmalsausprägung betrachtet.

Als erklärende Variablen wurden die Variablen Kulturart (Abkürzung ‚Grp-Veg‘), Vorfrucht (Abkürzung ‚Grp-Vor‘) und bewirtschaftender Betrieb (Abkürzung ‚Grp-Betr‘) verwendet. Als Varianten wurden die Schätzungen mit (Appendix ‚-CV‘ an Verfahrenskürzel) und ohne Kreuzvalidierung (Appendix ‚-RE‘ an Verfahrenskürzel; letzteres entspricht der Resubstitutionsschätzung der *classification and regression trees*) untersucht. Das Verfahren ist als Spezialfall eines Regressionsbaumes mit einer einzigen Variable anzusehen. Die Stichprobe der Beprobung vom Februar 1992 setzt sich aus jeweils 18 bewirtschaftenden Betrieben, 8 aktuellen Kulturarten und 12 Vorfrüchten zusammen.

3.6 Schätzung der Korngrößenfraktionsanteile

Die Schätzung der Korngrößenfraktionsanteilen durch das Weiblen-Verfahren ist in Kapitel 2.3.7 beschrieben. Die Ergebnisse von Weiblen (1997) stellen die Ausgangsgrundlage der vorliegenden Arbeit dar. Nachfolgend wird die Schätzung der Korngrößenfraktionsanteilen mit Hilfe der *classification and regression trees* kurz beschrieben.

3.6.1 Schätzung der Korngrößenfraktionsanteile mit CART

Mit Hilfe der 441 Horizontpaare Bodenschätzung – bodenkundliche Kartierung wurden mit CART Version 3.1 von Salford Systems für jede Korngrößenfraktion (Sand, Ton, Schluff) der bodenkundlichen Kartieranleitung Regressionsbäume zur Bestimmung der Korngrößenfraktionsanteile erstellt. Hierbei wurden die Zusatzinformationen sowohl als nominalskalierte (Bodenart nach Bodenschätzung) als auch stetige Variablen (Bodenzahl, Ackerzahl und Hangneigungsabschlag der Bodenschätzung) eingelesen und verarbeitet.

Da aufgrund des programmäßig implementierten CART-Algorithmus die Kreuzvalidierung nur bei der Baumerstellung, bzw. beim Baumschnitt realisiert ist, kann die Kreuzvalidierung nur für die Ableitung des Regressionsbaums ohne größeren manuellen Aufwand durchgeführt werden. Das Ergebnis der Ableitung ist zwar ein kreuzvalidierter Baum, es ist aber nicht möglich, alle unabhängig ermittelten Einzelergebnisse in einer Datei automatisiert zusammenzufassen. Eine derartige Zusammenfassung macht eine manuell durchgeführte Kreuzvalidierung erforderlich, bei der sukzessive ein Datenelement aus der gesamten Datenmenge ausgeblendet, danach ein Regressionsbaum erstellt und das Ergebnis für das Datenelement manuell in eine Datei geschrieben wird. Bei einem Datenumfang von 441 Horizontpaaren bei jeweils drei Zielvariablen war diese Lösung für die Ableitung der Bodenart nicht praktikabel. Aus diesem Grunde konnten nur die Ergebnisse der Resubstitutionsschätzung in einer Datei festgehalten und zur Ermittlung der Güte ausgewertet werden.

4 Fallstudie 1:

Schätzung von Nitratwerten im Boden

4.1 Ergebnisse der CART-Schätzungen

Die Resubstitutions-Regressionsbäume zur Schätzung der Nitratwerte sind in den Abbildungen 4.1-4.3 für die drei Tiefenstufen 0-3 dm, 3-6 dm und 6-9 dm dargestellt. Die entsprechenden CART Ausgabeprotokolle sind auf der beigelegten CD-ROM enthalten („CART Ausgabeprotokoll Nitrat.doc“). Die drei Ausgabeprotokolle enthalten jeweils am Ende auch die Parametervoreinstellungen auf deren Darstellung an dieser Stelle verzichtet wird. Auf eine Darstellung der 48 Kreuzvalidierungs-Regressionsbäume wird ebenfalls verzichtet. Die Einzelergebnisse für jedes Datenelement sind für alle in Kap. 3.4 und 3.5 erläuterten Verfahren und Varianten (Resubstitutionsschätzung und echte Kreuzvalidierungs-Schätzung) als MS-EXCEL Arbeitsmappe auf der beigelegten CD-ROM enthalten („Ergebnisse Nitratgehalt im Boden.xls“). Die Arbeitsmappe gliedert sich in mehrere Tabellenblätter, welche jeweils die Ergebnisse einer Variante eines Schätzverfahrens beinhalten.

4.2 Vergleich der Ergebnisse

Die Ergebnisse der Schätzverfahren und ihrer Varianten sind für jede Tiefenstufe in Abb. 4.1 zusammenfassend in Form von Schätzwert vs. Messwert – Diagrammen dargestellt. Die Ergebnisse der in Kap. 3.2 vorgeschlagenen Gütekriterien mit einem Huber-k-Schätzer zur Berechnung des robusten Mittelwertes (Mit k = Median der Medianabweichung = MAD) werden anschließend vorgestellt und diskutiert.

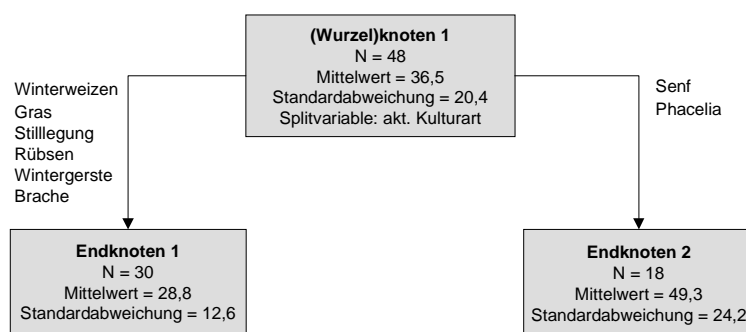


Abb. 4.1: Regressionsbaum zur Schätzung des Nitratgehaltes in der Tiefe 0-3 dm

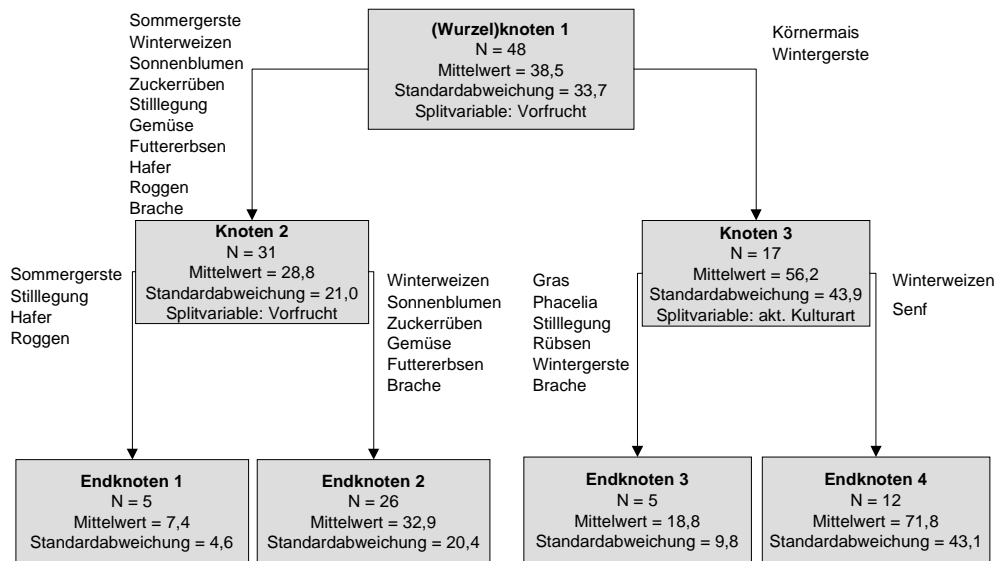


Abb. 4.2: Regressionsbaum zur Schätzung des Nitratgehaltes in der Tiefe 3-6 dm

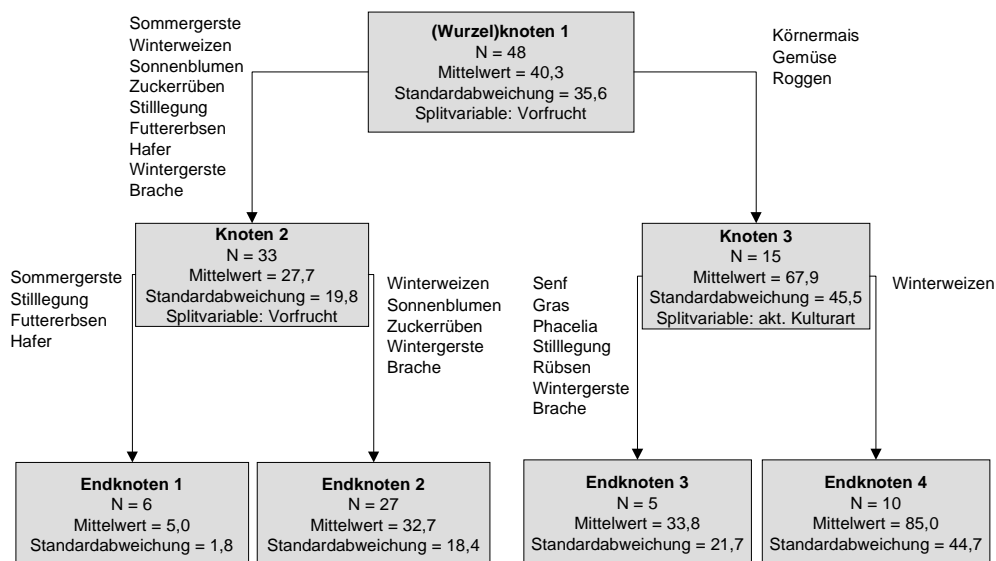


Abb. 4.3: Regressionsbaum zur Schätzung des Nitratgehaltes in der Tiefe 6-9 dm

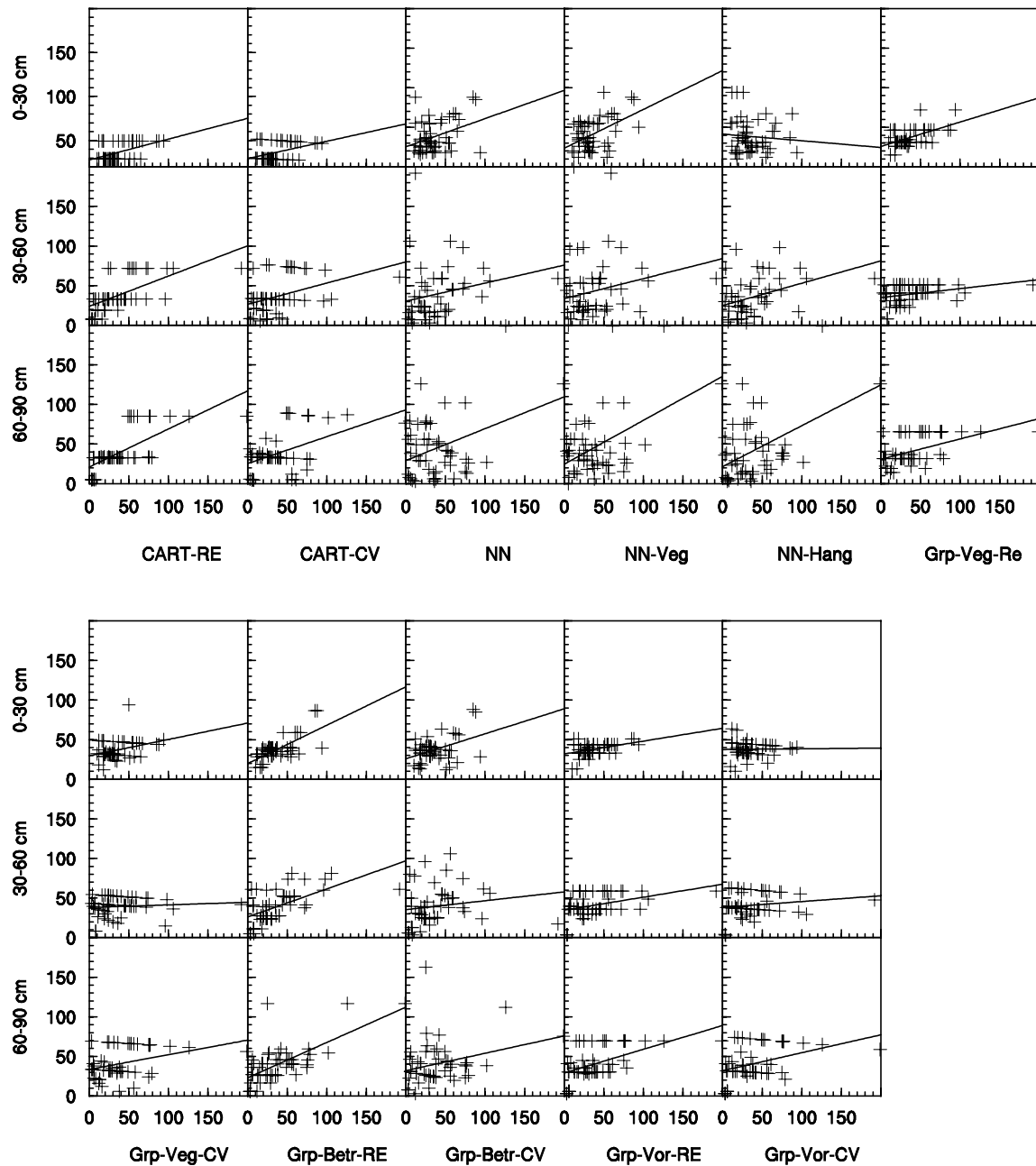


Abb. 4.4: Schätzwerte vs. Messwerte der Verfahren zur Schätzung von Nitratgehaltes

Ordinaten und Abszisseneinheiten in $\text{kg}/(\text{ha} \cdot 30 \text{ cm})$. Auf der Ordinate sind die Schätzwerte, auf der Abszisse die Messwerte der angegebenen Tiefenstufen und der betreffenden Schätzverfahren dargestellt. Zusätzlich sind die (kleinste-Quadrate) Ausgleichsgeraden eingezeichnet. Die Parameter der Regressionsgerade sind in den Abb. 4.12 und 4.13 vergleichend dargestellt.

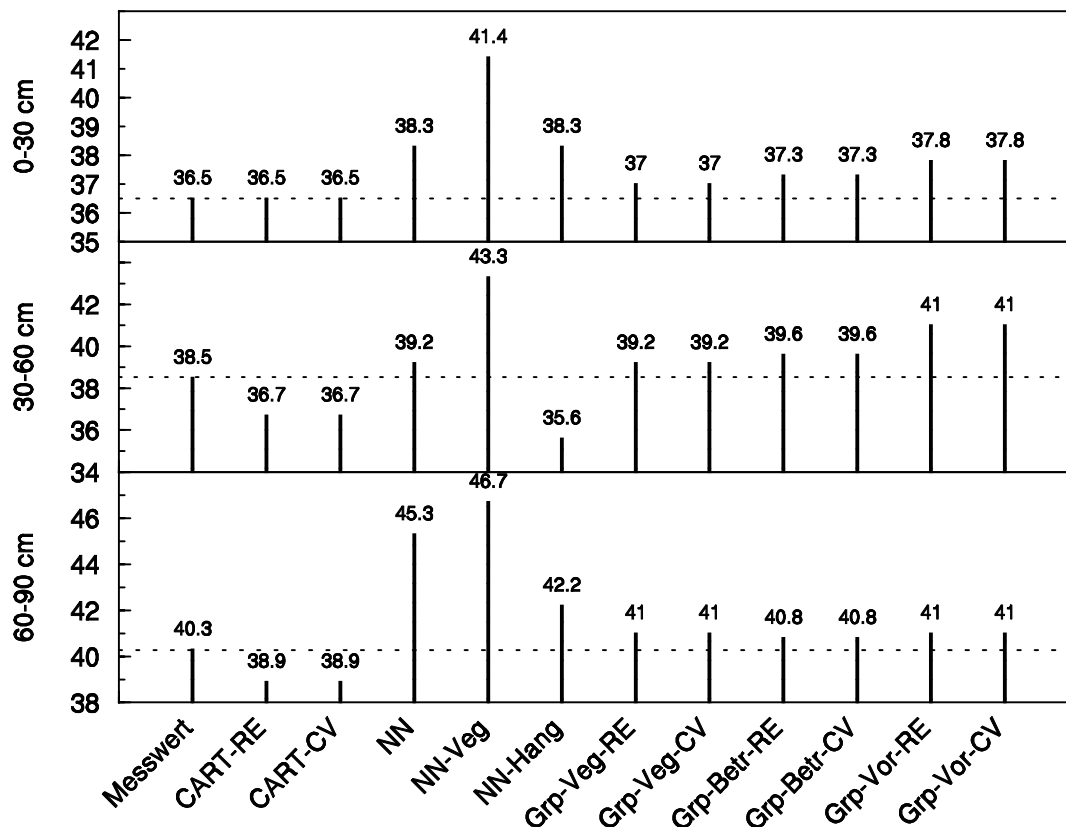


Abb. 4.5: Arithmetisches Mittel der Mess- und Schätzwerte nach Tiefenstufe geordnet.

*Gepunktete Linien kennzeichnen die Messwerte. Ordinateneinheiten in kg/(30 cm*ha)*

Die Mittelwerte der Tiefenstufe 0-30 cm sind gekennzeichnet durch eine Übereinstimmung bei beiden CART-Varianten mit dem Mittelwert der Messwerte und geringe Abweichungen (bis 4 %) der Gruppenmittelwertverfahren. Alle drei Nächste-Nachbarn Verfahren weisen höhere Abweichungen (bis maximal 13 % bei NN Veg) auf. Das CART Verfahren zeigt im vorliegenden Falle die beste, die Schätzung Nächster-Nachbar-gleicher-Vegetation hingegen die geringste Übereinstimmung zwischen dem Mittelwert der Messwerte und dem Mittelwert der Schätzwerte.

Auch bei der Tiefenstufe 30-60 cm gibt es nur relativ geringe Abweichungen vom Mittelwert der Messwerte: Bei den CART Varianten um knapp 5 %, bei der einfachen Nächste-Nachbar-Schätzung und die Gruppenmittelwertverfahren Grp-Veg und Grp-Betr um 3 %. Lediglich NN-Veg, NN-Hang und Grp-Vor weisen deutliche Abweichungen von über 5 % auf.

Auch in der Tiefenstufe 60-90 cm weichen die CART-Mittelwerte vom Mittelwert der Messwerte ab, die Gruppenmittelwertverfahren weisen alle kleinere Abweichungen als die CART Varianten auf. Die verschiedenen Gruppenmittelwertverfahren unterscheiden sich nur unwesentlich voneinander; die Nächste-Nachbar-Schätzer weisen vergleichsweise hohe Abweichungen bis 15 % auf.

In allen drei Tiefenstufen stimmen die kreuzvalidierten mit den nichtkreuzvalidierten Verfahren überein. Auffällig ist, dass die Gruppenmittelwertverfahren bei allen Tiefenstufen und Varianten immer einen höheren Mittelwert aufweisen als die Messwerte, d. h. die Messwerte werden eher überschätzt; bei den CART Verfahren ist jedoch der gegenteilige Befund festzustellen. Das CART Verfahren weist in der Tiefenstufe 0-30 cm die beste Übereinstimmung der arithmetischen Mittel der Schätzungen und

der Messungen auf, in den Tiefenstufen 30-60 cm und 60-90 cm haben die Gruppenmittelwertverfahren die kleinsten Abweichungen.

Die Nächste-Nachbarn Schätzverfahren NN und NN-Veg scheinen insbesondere mit zunehmender Tiefe grundsätzlich wenig geeignet zu sein, den Mittelwert der Messwerte abzubilden. Die Variante NN-Hang nimmt diesbezüglich in allen drei Tiefenstufen eine Mittelposition ein.

CART und die Gruppenmittelwertverfahren weisen einen Bias bis 6 % der Mittelwerte der Messwerte auf. Im Vergleich zu den Standardabweichungen der Messwerte (siehe Abb. 4.7) ist dieser Wert als gering zu bewerten. Die genannten Schätzverfahren sind daher hinsichtlich des Bias als brauchbar einzustufen.

Die Mittelwerte können von Extremwerten stark beeinflusst sein, demzufolge sind selbst bei ‚guten Schätzungen‘ Abweichungen vom Mittelwert der Messwerte möglich. Die Interpretation der Mittelwerte als ‚brauchbare‘ oder gar ‚gute‘ Werte zur Gebietscharakterisierung gestaltet sich daher mit den nichtrobusten Kennwerten sehr schwierig.

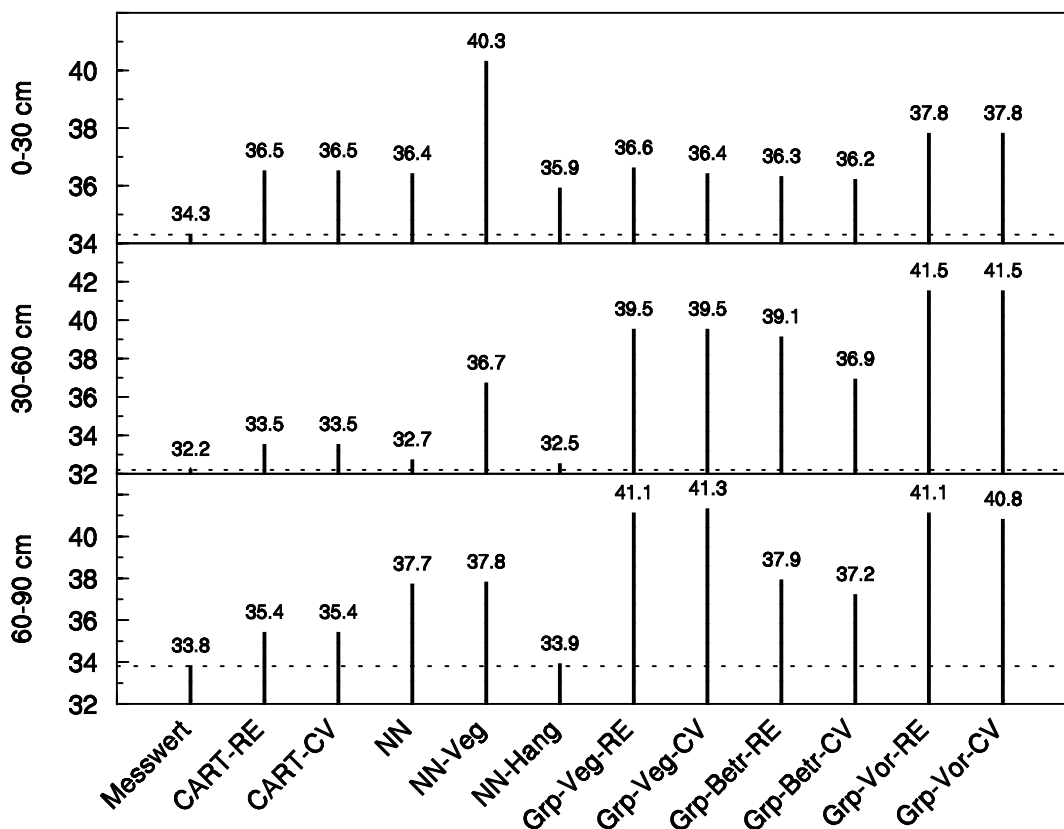


Abb. 4.6: Robuste Mittelwerte der Mess- und Schätzwerte nach Tiefenstufe geordnet

Gepunktete Linien kennzeichnen die Messwerte. Ordinateneinheiten in $\text{kg}/(30 \text{ cm} \cdot \text{ha})$

Die robusten Mittel der Messwerte sind in den drei untersuchten Tiefenstufen durch eine unterschiedlich stark ausgeprägte Abnahme gegenüber den arithmetischen Mitteln der Messwerte gekennzeichnet. In der Tiefenstufe 0-30 cm beträgt die Abnahme 2,2, in den folgenden Tiefenstufen 6,3 bzw. 6,5 $\text{kg}/(30 \text{ cm} \cdot \text{ha})$. Die deutliche Abnahme der robusten Mittel der Messwerte in den Tiefenstufen 30-90 cm ist auf eine Reihe hoher Messwerte (mehrere Werte über $100 \text{ kg}/(30 \text{ cm} \cdot \text{ha})$) zurückzuführen.

(siehe Anhang 7). Es zeigt sich, dass im Gegensatz zum arithmetischen Mittel die robusten Mittel der Messwerte in allen drei Tiefenstufen annähernd gleich sind. Die Zunahme der arithmetischen Mittelwerte der Nitratwerte mit der Beprobungstiefe ist daher nur auf den Einfluss einiger extremer Werte in den unteren Tiefenstufen zurückzuführen. Extreme Werte in den unteren Tiefenstufen können mit Problemkulturen (z. B. Mais) in vorhergehenden Vegetationsperioden und zwischenzeitlich erfolgter Stickstoffverlagerung erklärt werden.

Bei fehlerbehafteten Daten sind robuste Mittelwerte demnach besser als arithmetische Mittelwerte geeignet, die tendenzielle Lage der Messwerte zu beschreiben. Robuste Verfahren ergänzen daher sinnvoll nichtrobuste Verfahren.

In der obersten Tiefenstufe ist bei den robusten Mittelwerten der CART-Verfahren keine Abnahme im Vergleich zu den arithmetischen Mittelwerten zu verzeichnen. Bei den übrigen Verfahren ist die Abnahme meistens gering (maximal 6 % bei NN-Hang). In mittlerer Tiefe findet sich eine Abnahme der robusten Mittel bei CART und dem NN-Hang Verfahren und sogar leichte Zunahmen bei den Gruppenmittelwertverfahren im Vergleich zu den nichtrobusten Mittelwerten. In der Tiefenstufe 60-90 cm können vergleichsweise starke Abnahmen der robusten Mittel bei CART und NN-Veg (19 %) festgestellt werden, wohingegen bei den Gruppenmittelwertverfahren nur geringe Veränderungen vorliegen (Abnahme um maximal 9 %). Dieser Befund deutet darauf hin, dass NN-Hang und die CART-Verfahren unter den untersuchten Schätzverfahren am besten geeignet sind, tendenziell niedrige Messwerte, die von extremen Werten beeinflusst werden, nachzubilden. Die Gruppenmittelwertverfahren hingegen führen auf Grund ihrer größeren Gruppenzahl (mehr als vier Gruppen) zu einem höheren robusten Mittelwert als das CART Verfahren (höchstens vier Gruppen).

Wie schon beim arithmetischen Mittel, so stimmen auch bei den robusten Mittelwerten die kreuzvalidierten mit den nicht kreuzvalidierten Verfahrensvarianten mit Ausnahme von Grp-Betr im wesentlichen überein, d. h. die Schätzungen können bezüglich des arithmetischen und robusten Mittelwertes als stabil angesehen werden. Damit scheint eine ausreichende Stichprobenanzahl für eine verlässliche Schätzung zumindest für das untersuchte Datenensemble im Untersuchungsgebiet gegeben zu sein.

Robuste Verfahren sind insbesondere zur Qualitätskontrolle von administrativen Programmen geeignet, wenn von fehlerhaften Messwerten ausgegangen werden muss z. B. bei noch nicht ausgereiften bzw. fehleranfälligen Untersuchungsverfahren.

In Hinsicht auf eine robuste Lageschätzung ist im vorliegenden Fall das NN-Hang-Verfahren in allen drei Tiefenstufen mit einer maximalen Abweichung von 5 % (in der Tiefenstufe 0-30 cm) als das beste der untersuchten Verfahren zu charakterisieren.

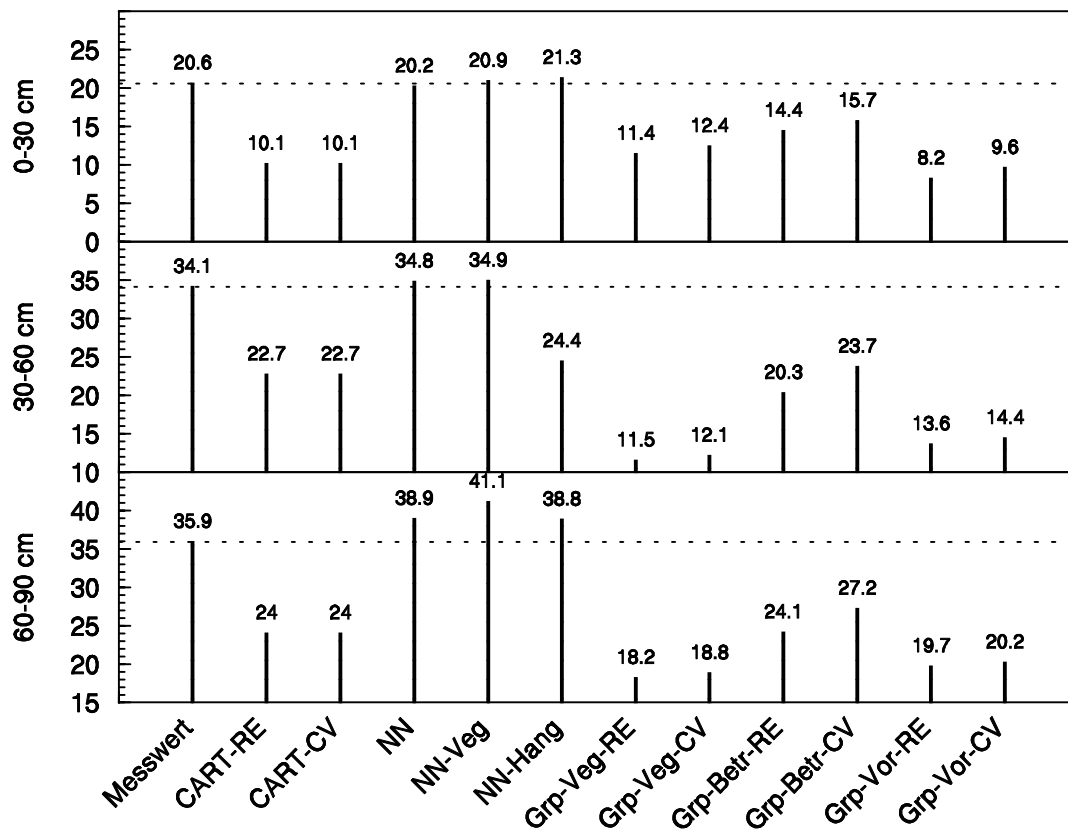


Abb. 4.7: Standardabweichungen der Mess- und Schätzwerte nach Tiefenstufe geordnet

*Gepunktete Linien kennzeichnen die Messwerte. Ordinateneinheiten in kg/(30 cm*ha)*

Die Standardabweichung der Mess- und Schätzwerte gibt Aufschluss über die Homogenität der Schätzverfahren im Vergleich mit den Messwerten. Die Standardabweichung der Messwerte nimmt von der obersten zur zweiten deutlich und zur dritten Tiefenstufe leicht zu. Die Ursache liegt in den Extremwerten (s.o.), die im Untersuchungsgebiet sporadisch auftreten. Aufgrund der Nitratauswaschung aus der ersten Tiefenstufe ist ihr Vorkommen auf die Tiefenstufen 30-90 cm begrenzt. Aufgrund ihres mittelnden Verfahrens sind die Schätzwerte von CART und den Gruppenmittelwertverfahren sehr viel homogener als die Messwerte, daher weisen sie generell geringere Standardabweichungen auf. Die NN-Verfahren haben hingegen eine ähnliche Standardabweichung wie die Messwerte. Bei den kreuzvalidierten Varianten der Gruppenmittelwertverfahren sind die Standardabweichungen größer als die nichtkreuzvalidierten Varianten. Bei CART tritt dieser Effekt nicht auf, die Regressionsbäume sind bezüglich der Standardabweichung stabil.

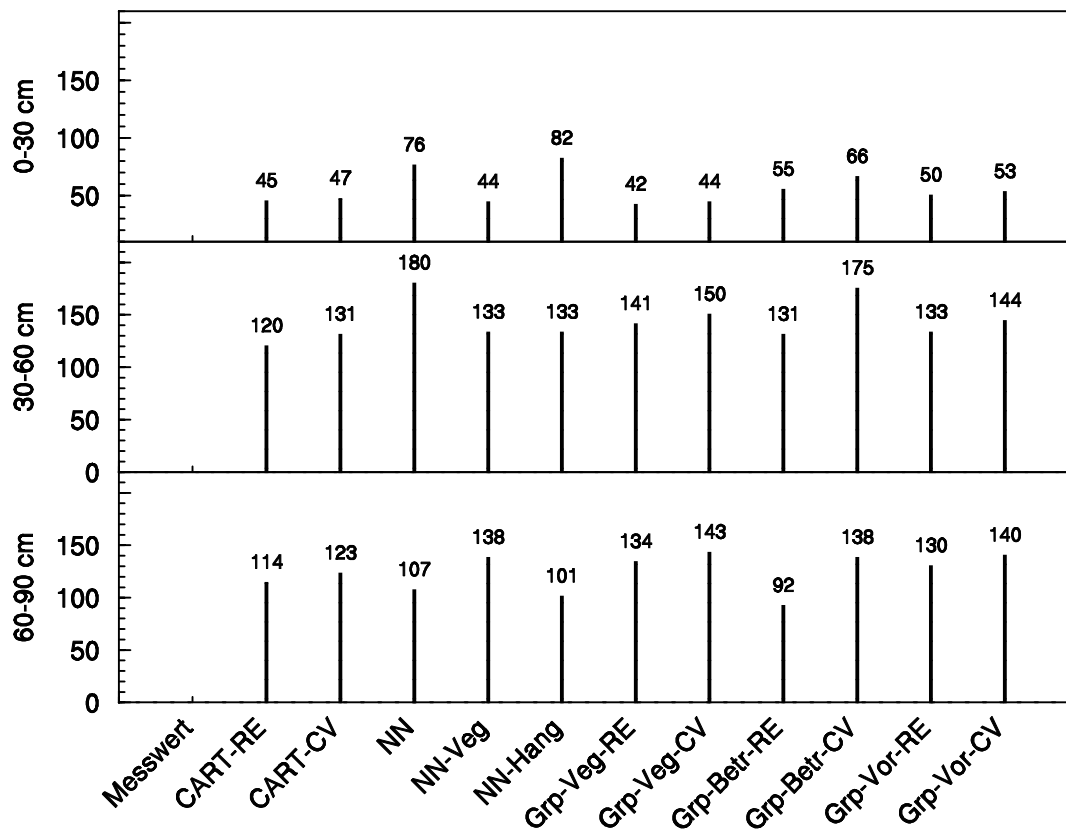


Abb. 4.8: Maximale Schätzabweichung (me) der Schätzwerte nach Tiefenstufe geordnet

Ordinateneinheiten in $\text{kg}/(30 \text{ cm} \cdot \text{ha})$

Bei der maximalen Schätzabweichung (me) findet sich ein großer Unterschied zwischen der obersten und den weiteren Tiefenstufen: In der ersten Tiefenstufe schwankt sie zwischen 44 und 82 $\text{kg}/(30 \text{ cm} \cdot \text{ha})$, in der mittleren und untersten Tiefenstufe differiert sie zwischen 92 und 180 $\text{kg}/(30 \text{ cm} \cdot \text{ha})$.

Alle kreuzvalidierten Varianten (auch beim CART-Verfahren) weisen eine größere maximale Schätzabweichung auf, als die nichtkreuzvalidierten Varianten. Auch hier liegt die Ursache darin, dass eine Reihe hoher Messwerte (mehrere Werte über 100 $\text{kg}/(30 \text{ cm} \cdot \text{ha})$) vorliegt (s. o.), die bei der Resubstitutionsschätzung den Mittelwert der jeweiligen Gruppe anhebt und somit zu kleineren Schätzabweichungen als bei der Kreuzvalidierungsschätzung führt.

Bei Betrachtung der kreuzvalidierten Varianten der Schätzverfahren ist das Bild in den vorliegenden Tiefenstufen bezüglich der maximalen Schätzabweichung uneinheitlich: In der ersten Tiefenstufe haben NN-Veg und Grp-Veg-CV die kleinste me , in der zweiten Tiefenstufe CART-CV und in der dritten Tiefenstufe das Verfahren NN-Hang, welches in der Tiefenstufe 0-30 cm jedoch die größte me hat. CART vermag also im vorliegenden Falle in den ersten beiden Tiefenstufen große Werte mit kleineren Schätzabweichungen nachzubilden, als die übrigen Verfahren. Trotz dieser Eigenschaft sind sowohl das arithmetische als auch das robuste Mittel (s. Abb. 4.6 und 4.6) gleich oder kleiner als die der Vergleichsverfahren. Dies belegt die Vermutung, dass CART bei gleicher Datenbasis differenziertere Schätzungen ermöglicht und vorhandene Informationen besser nutzen kann als die untersuchten Vergleichsverfahren.

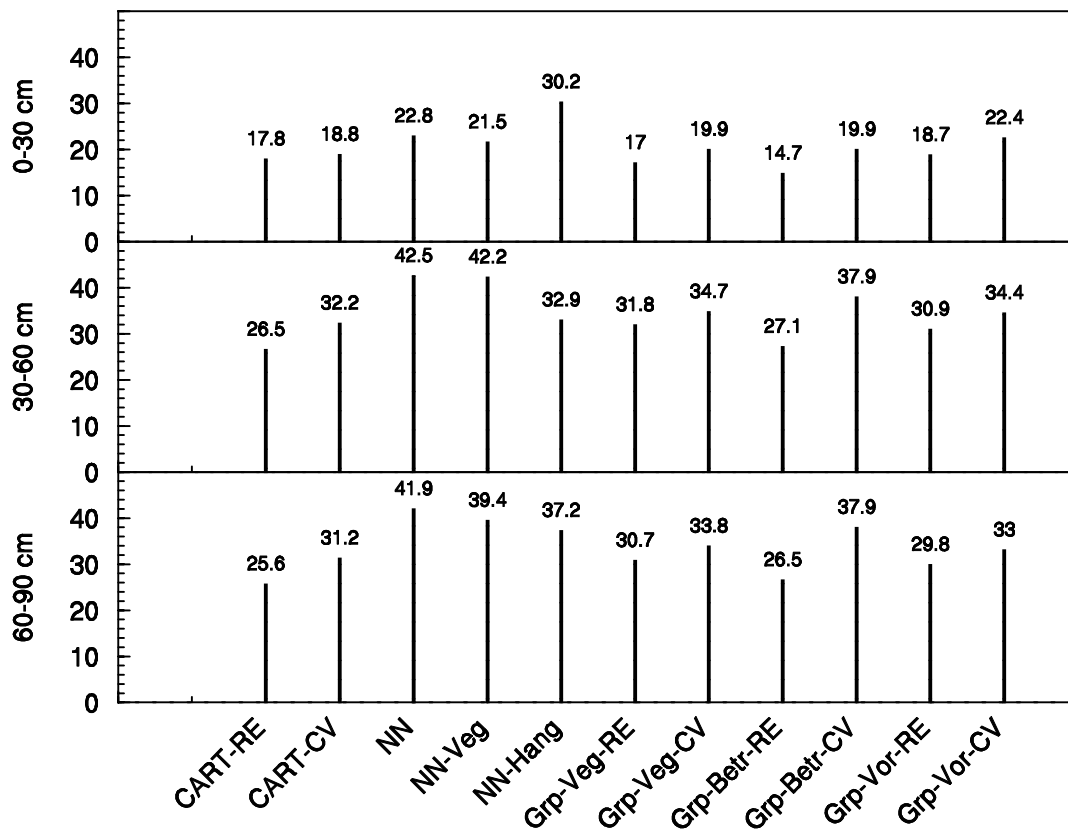


Abb. 4.9: Wurzel der mittleren quadrierten Schätzfehler (*rmsr*) der Schätzverfahren

Ordinateneinheiten in $\text{kg}/(30 \text{ cm} \cdot \text{ha})$

Die Wurzel der mittleren quadrierten Schätzabweichung („Standardisierter Schätzabweichung“) gibt Aufschluss über die Absolutbeträge der Residuen.

Die standardisierten Schätzabweichungen sind in der ersten Tiefenstufe allgemein geringer als in den Tiefenstufen 30-90 cm. In Bezug auf das arithmetische oder robuste Mittel der Messwerte (siehe Abb. 4.5, 4.6) sind die standardisierten Schätzabweichungen als hoch einzustufen.

Auch bei den standardisierten Schätzabweichungen zeigt sich, dass die kreuzvalidierten Varianten in allen Tiefenstufen deutlich größer sind als die nichtkreuzvalidierten Varianten und daher die Ergebnisse der Schätzungen bei der vorliegenden Datenlage als wenig stabil angesehen werden müssen.

Unter den kreuzvalidierten Varianten (zu denen sinngemäß auch die NN-Verfahren dazugezählt werden) hat das CART-Verfahren in allen drei Tiefenstufen die geringste standardisierten Schätzabweichung.

Diese Befunde belegen erneut die Vermutung, dass das CART Verfahren eine besser angepasste Schätzung ermöglicht als die untersuchten Vergleichsverfahren.

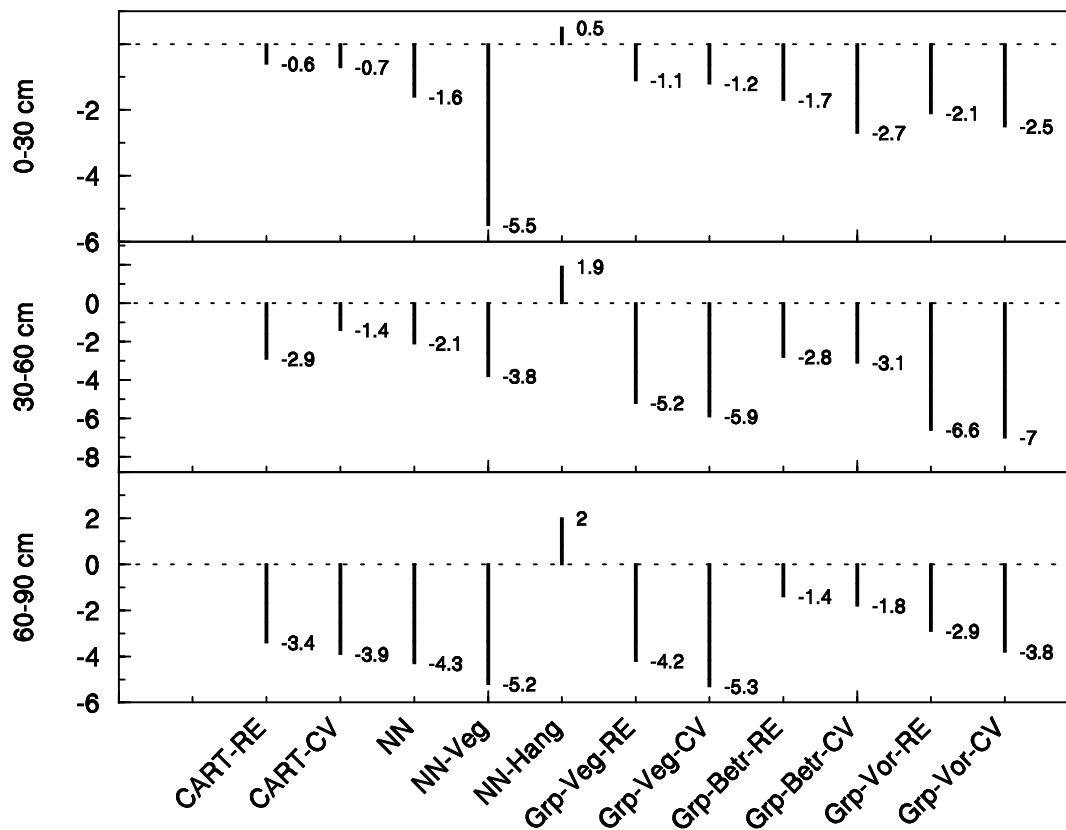


Abb. 4.10: Robustes Mittel (*rer*) der Schätzabweichungen der Schätzverfahren

*Ordinateneinheiten in kg/(30 cm*ha)*

Das robuste Mittel der Schätzabweichungen ist in allen Tiefen mit Ausnahme von NN-Hang negativ, da sehr große Schätzabweichungen aufgrund extremer Messwerte durch die robuste Mittelung gestutzt werden. Die robuste Lage der Schätzwerte ist daher höher als die der Messwerte.

Der Betrag von *rer* liegt bei den besten kreuzvalidierten Varianten von CART und den Gruppenmittelwertverfahren zwischen 0,7 bis 1,8 kg/(30 cm*ha) und damit im Bereich zwischen 2 und 5 % des robusten Messwertes. Die robusten Mittel der Schätzabweichungen sind daher als relativ gering anzusehen. Im Vergleich mit der Wurzel der mittleren quadrierten Schätzabweichungen ist *rer* eine Größenordnung kleiner. Die Ursache dieser unterschiedlichen Befunde liegt bei großen Schätzabweichungen, die die Wurzel der mittleren quadrierten Schätzabweichungen sehr stark dominieren.

NN-Hang erweist sich in den zwei Tiefenstufen 0-30 cm und 60-90 cm als das Schätzverfahren mit dem kleinsten mittleren robusten Schätzabweichung; CART nimmt diese Funktion in der Tiefenstufe 30-60 cm ein und hat auch in der Tiefenstufe 0-30 cm nur eine unwesentlich größere *rer* als NN-Hang. Die kreuzvalidierten Varianten haben mit Ausnahme von CART in der Tiefe von 30-60 cm eine größere *rer* als die Resubstitutionsschätzungen, d. h. die Schätzergebnisse sind in Hinsicht auf den Datensatz nicht stabil.

In den beiden Tiefenstufen 30-60 cm und 60-90 cm weisen beim vorliegenden Datensatz CART bzw. NN-Hang sowohl die kleinste *rer* als auch die kleinste maximale Schätzabweichung auf.

Interessanterweise hat NN-Hang in der Tiefe 0-30 cm jedoch die kleinste *rer*, aber die größte maximale Schätzabweichung, d. h. das Verfahren liefert grundsätzlich Schätzungen mit geringen Schätzabweichungen, hat aber gleichzeitig die größte maximale Schätzabweichung (*me*) aller Verfahren.

Aus diesen unterschiedlichen Befunden für die oberste und die darunter liegenden Tiefenstufen lässt sich die Vermutung ableiten, dass die Entfernung und die Hanglage wichtige erklärenden Variable des Stickstoffgehaltes darstellen, und dass ein einzelner hoher Messwert in der Umgebung von sehr niedrigen Messwerten (siehe Abb. 4.4) bei gleicher Hanglage eine maximale Schätzabweichung verursacht, die größer ist als bei den Vergleichsverfahren. Dies kann als Hinweis darauf verstanden werden, dass das NN-Hang-Verfahren weniger robust ist als die übrigen Verfahren.

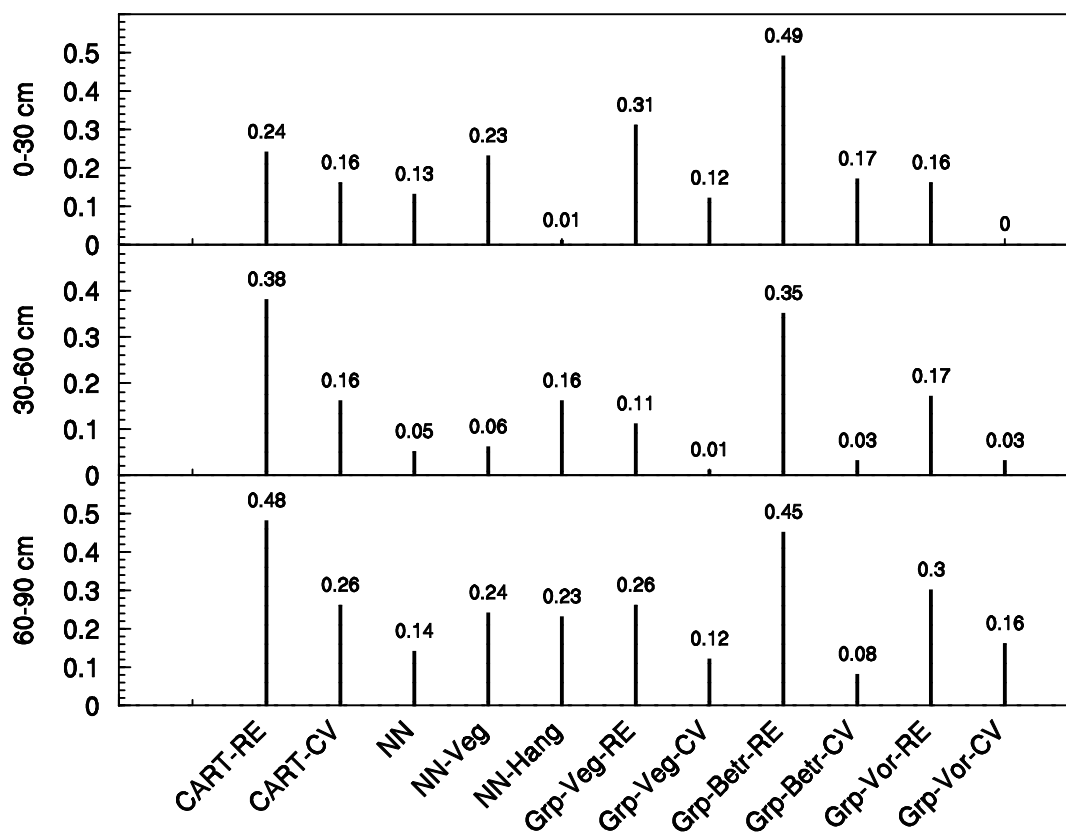


Abb. 4.11: Bestimmtheitsmaß (*cdr*) der Ausgleichsgeraden zwischen gemessenen und geschätzten Werten

Das Bestimmtheitsmaß der Regressionsgeraden zwischen den gemessenen und den geschätzten Werten schwankt von Null (Tiefenstufe 0-30 cm, Grp-Vor-CV) bis maximal 0,49 (Tiefenstufe 0-30 cm, Grp-Betr-RE). Die kreuzvalidierten Varianten weisen in allen Tiefenstufen und bei allen Verfahren ein deutlich geringeres Bestimmtheitsmaß auf als die nicht kreuzvalidierten Varianten. Im Vergleich mit den Gruppenmittelwertverfahren ist die relative Abnahme bei den CART-Varianten jedoch noch am kleinsten. Die Regressionsgeraden sind aufgrund der geringen Bestimmtheitsmaße als wenig bestimmt anzusehen.

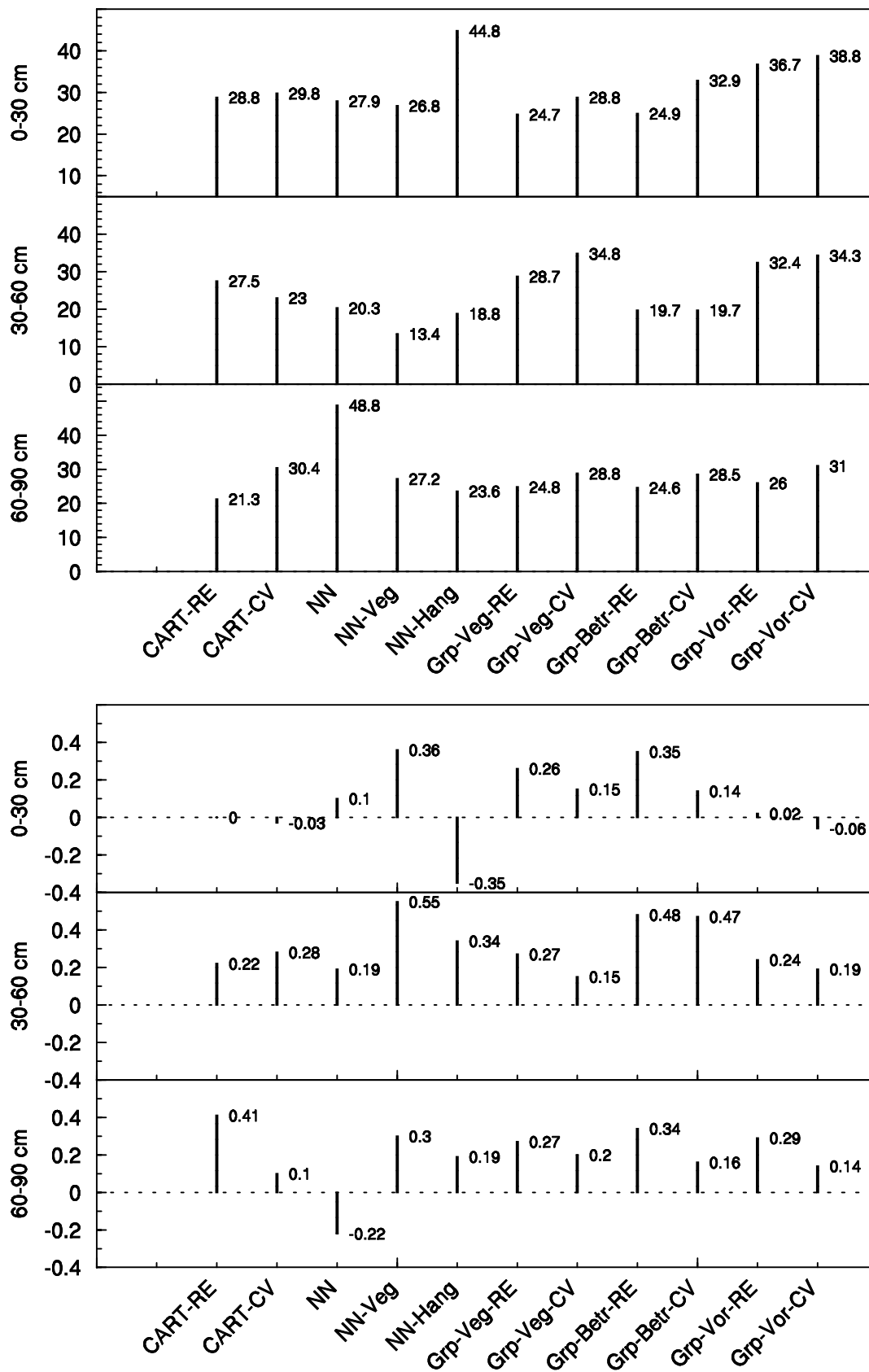


Abb. 4.12 und 4.13: Achsenabschnitt (a) und Steigung (b) der robusten Regressionsgeraden zwischen gemessenen und geschätzten Werten *Ordinateneinheiten des Achsenabschnittes in kg/(30 cm*ha)*

Die robuste Ausgleichsgerade nach Tukey (1977; vgl. Kap. 3.2.3.5) zwischen den gemessenen und den geschätzten Werten ist bei allen Verfahren und in allen Tiefen durch sehr hohe Achsenabschnitte (mindestens 68 % des arithmetischen Mittelwertes der Messwerte) und Steigungen deutlich geringer als 1 (bis 0,55; teilweise negative Steigungen) gekennzeichnet. Dies dokumentiert eine systematische Überschätzung bei kleinen Messwerten und eine Unterschätzung bei hohen Messwerten bei allen Verfahren. Die zwei Verfahren NN-Veg und Grp-Betr (Grp-Betr mit den beiden Varianten RE und CV) haben dabei die kleinsten Achsenabschnitte und relativ größten Steigungen.

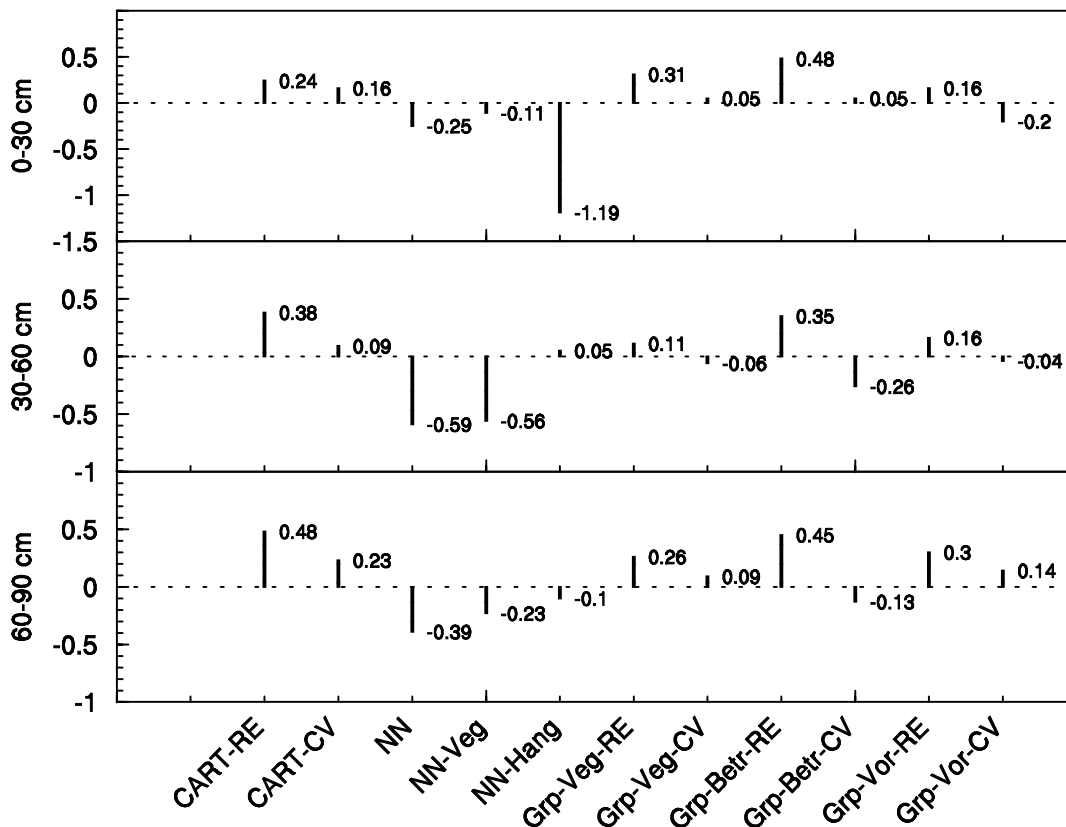


Abb. 4.14: Koeffizient der mittleren quadrierten Schätzabweichung (*comsre*) nach Tiefenstufe geordnet

Der bestmögliche Wert von *comsre* ist 1; negative Werte bedeuten, dass die mittleren quadrierte Schätzabweichungen bei Verwendung des arithmetischen Mittelwertes als Schätzverfahren geringer wären als bei Verwendung des betreffenden Verfahrens.

Bei den NN-Verfahren liegen mit einer Ausnahme (NN-Hang in der Tiefe 30-60 cm) stark negative *comsre*-Werte vor. Hinsichtlich der mittleren quadratischen Schätzabweichung sind fast alle Varianten des Nächste-Nachbarn Verfahrens schlechter als eine jeweilige Schätzung durch den arithmetischen Mittelwert.

Bei den Gruppenmittelwertverfahren ist bei den Resubstitutionsschätzungen eine deutliche Reduktion der Schätzvarianz (zwischen 11 und 48 %) gegenüber der Verwendung eines Referenzschätzverfahrens, in Falle von *comsre* des Mittelwertes der Messwerte festzustellen. Die *comsre*-Werte der Kreuzvalidierungsschätzungen sind im Vergleich zu den jeweiligen Resubstitutionsschätzungen jedoch deutlich geringer, teilweise liegen sogar negative *comsre*-Werte vor. Damit müssen die Gruppenmittel-

wertverfahren in Bezug auf das Gütekriterium *comsre* als instabil bezeichnet werden. Eine mögliche Ursache hierfür sind 'überangepasste' Resubstitutionsschätzungen, bei denen alleine auf Grund einer großen Zahl von erklärenden Variablen gute Anpassungen bei der Resubstitutionsschätzung möglich sind.

Auch die kreuzvalidierte CART-Variante weist in allen drei Tiefenstufen einen großen Unterschied zwischen der Resubstitutionsschätzung und der Kreuzvalidierungsschätzung auf. Die Kreuzvalidierungsschätzungen in den untersuchten Tiefenstufen haben im Gegensatz zu den Gruppenmittelwertverfahren kleine, aber deutlich positive *comsre*-Werte, so dass die kreuzvalidierte CART-Variante in allen drei Tiefenstufen die größten *comsre*-Werte vorweist. Die frappierenden Unterschiede zwischen den Resubstitutionsschätzungen und den kreuzvalidierten Schätzungen belegen deutlich die Notwendigkeit, zur Beurteilung der Schätzverfahren die kreuzvalidierten Schätzungen mit heranzuziehen.

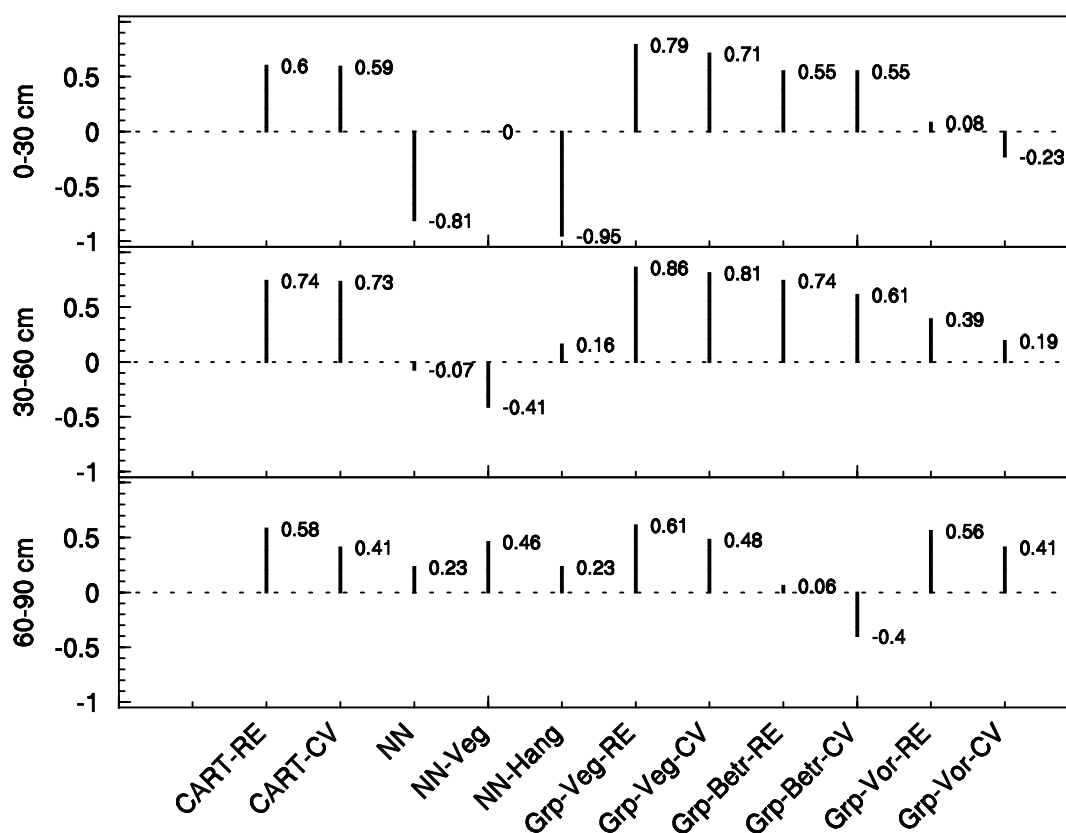


Abb. 4.15: Koeffizient der maximalen Schätzabweichung (*comare*) nach Tiefenstufe geordnet

Der bestmögliche Wert von *comare* ist 1; negative Werte bedeuten, dass eine schlechtere Schätzung vorliegt als bei Verwendung des arithmetischen Mittels aus minimalem und maximalem Messwert als Schätzverfahren.

Bei der maximalen Schätzabweichung (*comare*) finden sich allgemein deutlich größere Koeffizienten als bei *comsre*. Damit ist eine bessere Schätzung im Sinne einer relativ größeren Verringerung der spezifischen Schätzabweichung verbunden als bei *comsre*. Die *comare*-Werte der beiden CART Varianten Resubstitutionsschätzung und Kreuzvalidierungsschätzung unterscheiden sich in den ersten beiden Tiefenstufen um weniger als einen Prozentpunkt. Das CART-Verfahren ist daher hinsichtlich der maximalen Schätzabweichung in diesen Tiefenstufen als stabil zu bezeichnen. Die NN-Verfahren

haben nur in der dritten Tiefenstufe deutlich positive *comare*-Werte. In den ersten beiden Tiefenstufen führen die NN-Verfahren im allgemeinen zu größeren maximalen Schätzabweichungen als bei Verwendung eines konstanten Schätzwertes. Bei den Gruppenmittelwertverfahren sind ebenfalls – tiefen- und verfahrensabhängig – positive Werte mit über 50 % kleineren maximalen Schätzabweichungen festzustellen. Insgesamt sind die Unterschiede zwischen den Resubstitutions- und Kreuzvalidierungsvarianten allgemein kleiner als bei dem Gütekriterium *comsre*, d. h. die Schätzungen sind hinsichtlich des Gütekriteriums stabiler. Das Verfahren Grp-Veg hat in allen drei Tiefenstufen die größten *comare*-Werte, gefolgt von CART. Die maximalen Schätzabweichungen werden bei Grp-Veg je nach Tiefenstufe im Vergleich zu einer konstanten Schätzung sehr deutlich reduziert. Die relative Reduktion liegt zwischen knapp 50 und 80 %. Die Verfahren Grp-Veg und CART weist somit hinsichtlich des Gütemaßes *comare* die größten *core*-Werte unter allen Verfahren auf.

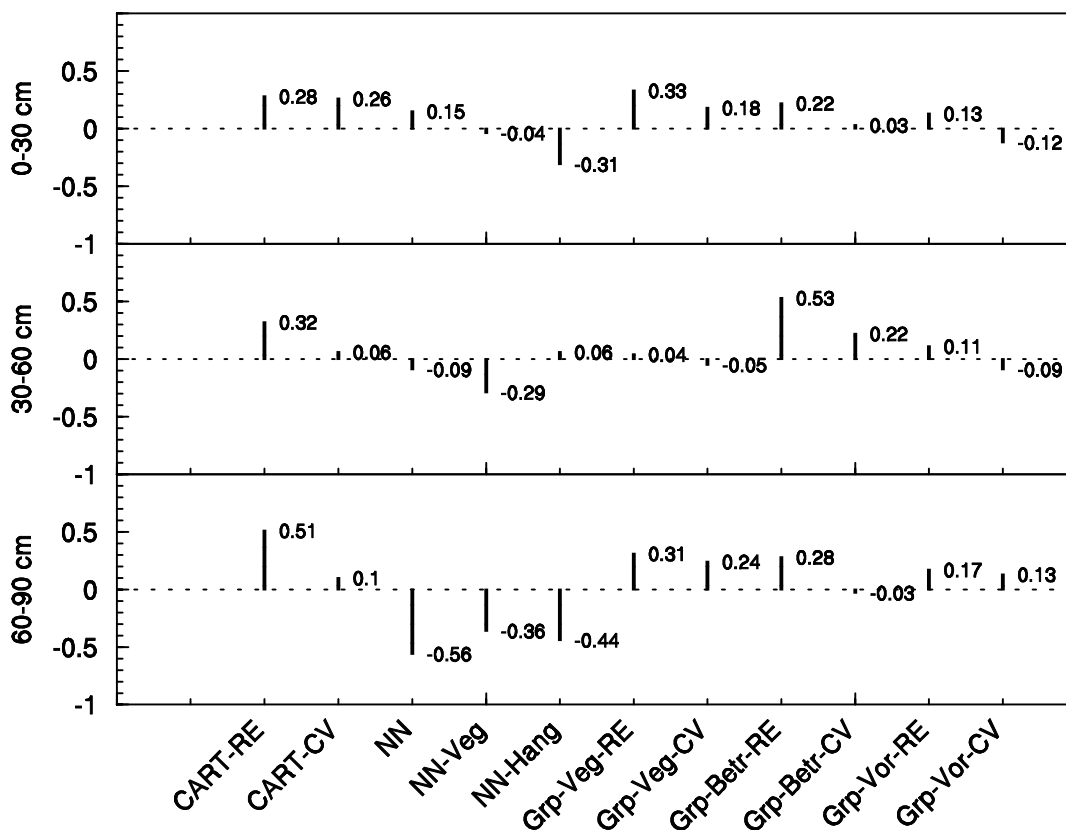


Abb. 4.16: Koeffizient des zentralen Schätzfehlers (*cocare*) nach Tiefenstufe geordnet

Der bestmögliche Wert von *cocare* ist 1; negative Werte bedeuten, dass eine schlechtere Schätzung vorliegt als durch Verwendung des Medians der Messwerte.

Bei den Gruppenmittelwertverfahren ist in Hinsicht auf eine Verringerung der zentralen Schätzabweichung in allen Tiefen ein deutlicher Unterschied zwischen den Resubstitutionsvarianten und den Kreuzvalidierungsvarianten, teilweise mit negativen *cocare*-Werten, festzustellen. Ebenso weisen die NN-Verfahren meist negative Koeffizienten auf.

In der ersten Tiefenstufe fällt der geringe Unterschied zwischen den beiden CART-Varianten auf. Bei den beiden CART Varianten ist ein vergleichsweise hoher *cocare*-Wert festzustellen. Im Gegensatz

dazu ist in den unteren Tiefenstufen der Unterschied zwischen den beiden Varianten groß. Das CART Verfahren ist hinsichtlich *cocare* daher nur in der ersten Tiefenstufe als stabil zu bezeichnen.

Bei den jeweils besten kreuzvalidierten Verfahren der drei Tiefenstufen sind die *cocare*-Werte deutlich größer als die entsprechenden *comsre*-Werte der gleichen Variante und Tiefe. Die zentrale Schätzabweichung wird durch die gleichen Schätzverfahren deutlicher reduziert als die mittlere quadrierte Schätzabweichung.

In jeder Tiefenstufe findet sich eine CV-Variante mit einer mindestens 20 %igen Verringerung des Medians der Schätzabweichung im Vergleich zum Median als konstanten Schätzwert. Eine entsprechende Reduzierung der zentralen Schätzabweichung durch ein geeignetes Schätzverfahren scheint daher in jeder Tiefenstufe möglich.

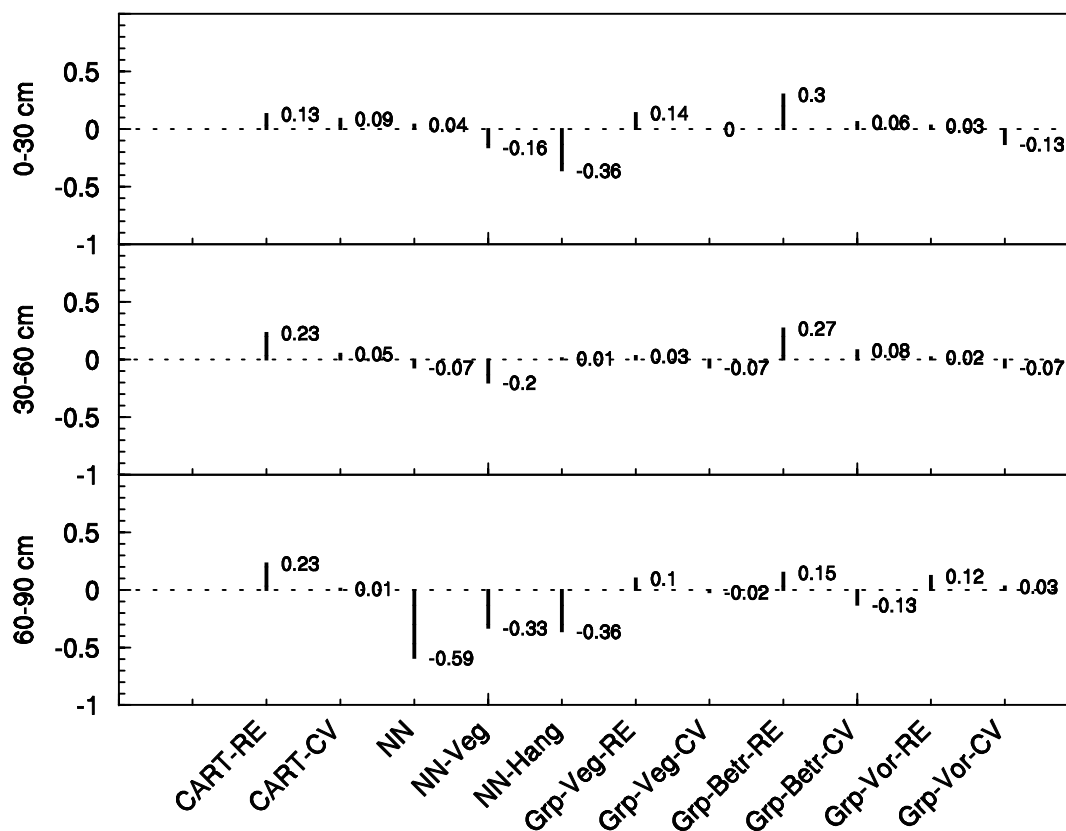


Abb. 4.17: Koeffizient des relativen robusten Schätzfehlers (*corore*) nach Tiefenstufe geordnet

Der bestmögliche Wert von *corore* ist 1; negative Werte bedeuten, dass eine schlechtere Schätzung vorliegt als durch Verwendung des robusten Mittelwertes als konstantes Schätzverfahren.

Die *corore*-Werte sind in allen drei Tiefenstufen grundsätzlich kleiner als die entsprechenden *comsre* bzw. *comare*-Werte der gleichen Tiefe und des gleichen Verfahrens. In den ersten beiden Tiefenstufen weisen die selben Varianten der Schätzverfahren die jeweils größten *corore*-, *comsre*- und *cocare*-Werte auf. Es gibt in den ersten beiden Tiefenstufen keine kreuzvalidierte Variante, die zu einer Verringerung der robusten Schätzabweichung um mehr als 9 % führt. In der dritten Tiefenstufe finden sich bei fast allen kreuzvalidierten Varianten nur *corore*-Werte kleiner oder gleich 0,03. Das robuste

Mittel als Schätzverfahren mit konstantem Schätzwert führt daher in den meisten Fällen zu einem geringeren robusten Mittelwert der Schätzabweichung als die untersuchten kreuzvalidierten Varianten der Schätzverfahren.

Betrachtet man die größten *core*-Werte der kreuzvalidierten Varianten einer Tiefenstufe, so lassen sich die Verfahren nach zunehmendem *core*-Wert wie folgt in eine Reihe bringen:

$$corore < comsre \approx cocare < comare$$

Corore hat unter den *core*-Kriterien den kleinsten Wert, was bedeutet, dass ein Standard, der durch *corore* dargestellt wird, schwerer zu erfüllen ist als durch die anderen untersuchten Kriterien. Es scheint das empfindlichste Maß für die Schätzabweichungen unter den dargestellten *core*-Kriterien zu sein. Aus diesem Grunde erscheint *corore* geeignet, als Mindeststandard-Gütemaß für Schätzungen verwendet zu werden. Dadurch ergibt sich ein interessantes Paradoxon: *corore* ist von den Annahmen der verwendeten mathematischen Mittel, weicher, der damit formulierbare Standard ist jedoch schwerer zu erfüllen als beispielsweise bei *comsre*.

4.3 Zusammenfassung der Ergebnisse

Bei fehlerbehafteten Daten ist das robuste Mittel per definitionem besser als der arithmetische Mittelwert geeignet, die allgemeine Lage der Messwerte zu beschreiben. Dies äußert sich bei dem untersuchten Datensatz darin, dass es große Unterschiede zwischen arithmetischem und robustem Mittel gibt, die auf eine Reihe großer Messwerte in den beiden unteren Tiefenstufen zurückzuführen sind.

Abhängig von der Anwendung der Schätzung können verschiedene Gütekriterien möglich und sinnvoll sein.

Es ist möglich, ein Set von *core*-Gütekriterien zu erstellen, um intrinsische Mindeststandards einer guten Schätzung festzulegen (mit *core*-Werten > 0); Eine allgemein gute Schätzung wäre z. B. eine Schätzung, bei der alle *core*-Werte größer Null sind.

Hinsichtlich der Unterschiede der Kreuzvalidierungs- und den Resubstitutionsvarianten sind nur beim arithmetischen und robusten Mittel geringe Unterschiede feststellbar. Alle anderen Gütemaße unterscheiden sich hinsichtlich der beiden Varianten wesentlich. Besonders auffallend ist der Unterschied beim Steigungsmaß der robusten Regression nach Tukey. Hier findet sich im Allgemeinen bei der CV-Variante nur eine halb so große Steigung wie bei der Resubstitutionschätzung, teilweise liegen sogar negative Steigungen vor. Die Verfahren bilden daher die Tendenz der Größendifferenzierung nur in geringem Ausmaß ab.

Abhängig von der Anwendung der Schätzung können unterschiedliche Schätzverfahren geeignet sein. Eine konstante Schätzung mit dem Nenner des *core*-Quotienten - $f(y_1 - a, \dots, y_N - a)$ - kann manchmal selbst ein geeignetes Schätzverfahren darstellen.

Keines der untersuchten Verfahren besticht in allen Tiefenstufen durch beste Gütewerte. Im Vergleich zu den NN und den Gruppenmittelwertverfahren zeichnet sich jedoch CART durch ausgewogene Ergebnisse bei allen Gütekriterien und in allen Tiefenstufen aus.

Teilweise sind die Gütemaße der Kreuzvalidierungsschätzungen bei CART und den Gruppenmittelwertverfahren erheblich kleiner als die der Resubstitutionsschätzungen. Dieser Effekt ist teilweise von Tiefenstufe zu Tiefenstufe unterschiedlich stark ausgeprägt. Die Ursache dieser wenig stabilen Schätzungen könnte eine zu geringe Stichprobenanzahl sein.

Um geringe Unterschiede zwischen der Kreuzvalidierungsschätzung und der Resubstitutionsschätzung zu erzielen, ist nach den vorliegenden Befunden die erforderliche Stichprobenanzahl von Gütemaß zu Gütemaß unterschiedlich.

Die jeweils größten *core*-Werte der kreuzvalidierten Schätzverfahren weisen eine große Spannweite auf. Der maximale Wert der Reduktion der relativen Schätzabweichung beträgt 81 % (Schätzverfahren Grp-Veg-CV in der Tiefenstufe 30-60 cm; Gütekriterium *comare*), der minimale 3 % (Schätzverfahren Grp-Vor-CV in der Tiefenstufe 60-90 cm; Gütekriterium *corore*). Die Güte der Schätzverfahren variiert damit je nach verwendetem Gütekriterium und Tiefenstufe von beachtlich bis marginal.

Das Güteniveau ist auch zwischen den drei Tiefenstufen unterschiedlich; es gibt keine Tiefenstufe, die sich gegenüber einer anderen hinsichtlich aller *core*-Werte auszeichnet.

5 Fallstudie 2:

Schätzung der Korngrößenfraktionsanteile

Im folgenden werden neben den Ergebnissen des Weiblen-Verfahrens die Resubstitutionsschätzungen der CART-Regressionsbäume für die Schätzung der Korngrößenfraktionsanteile vorgestellt. Die einzelnen Mess- und Schätzwerte der beiden Verfahren sind auf der beigelegten CD-ROM als MS-Excel Arbeitsmappe („Ergebnisse Bodenart.xls“) dokumentiert. In Abb. 5.1 ist exemplarisch der Regressionsbaum für die Schätzung des Tongehaltes abgebildet. Die Güte der Schätzungen auf Grundlage der in Kap. 3.2 vorgeschlagenen Gütekriterien mit einem Huber-k-Schätzer zur Berechnung des robusten Mittelwertes (Mit k = Medianabweichung = MAD) werden anschließend vorgestellt und diskutiert.

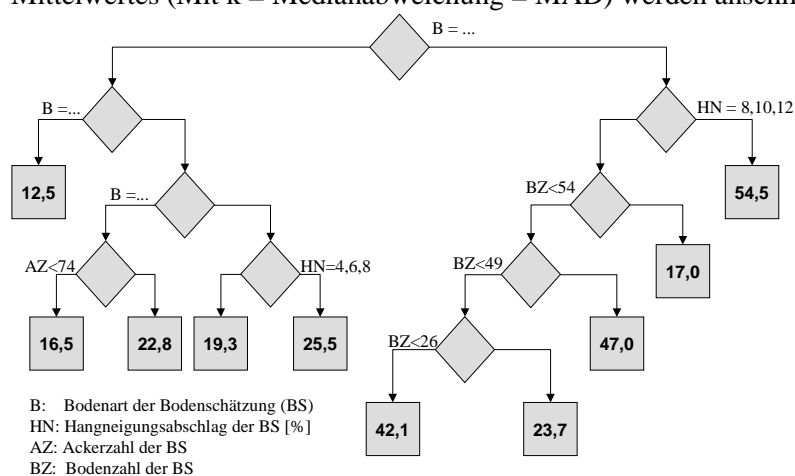


Abb. 5.1: Schematisierter CART-Regressionsbaum zur Schätzung des Tongehaltes

Exemplarisch dargestellt ist der Regressionsbaum zur Schätzung des Tongehaltes mit neun Knoten (10 Endknoten) und den betreffenden Knoten-Trennungsregeln. Als Vorhersagevariablen werden die Bodenart der Bodenschätzung (B; Kodierung der Bodenarten siehe Anhang 4: Numerische Kodierung der BS-Bodenarten), der Hangneigungsabschlag der Bodenschätzung (HN); Bodenzahl der Bodenschätzung (BZ), Abschlag auf Bodenzahl aufgrund Hangneigung [HN]=%, und die Ackerzahl der Bodenschätzung (AZ) verwendet. Die Trennregeln für den zweiten und dritten Knoten sind aufgrund ihres Umfangs nicht explizit dargestellt. An den Endknoten ist der Tongehalt in Masseprozenten angegeben. Er berechnet sich aus dem arithmetischen Mittel aller Elemente des Endknotens.

Die mit Hilfe des CART-Verfahrens selektierten Trennvariablen und die Zahl der Knoten der Regressionsbäume zur Schätzung der drei Korngrößenfraktionsanteile sind in Tabelle 5.1 abgebildet. Für jede Zielvariable wird vom CART-Verfahren ein unterschiedliches Set von Vorhersagevariablen verwendet. Für die Schätzung des Tongehaltes wurden von CART die BS-Bodenart, der Hangneigungsabschlag als Maß der Hangneigung, die Acker- und die Bodenzahl als relevante Vorhersagevariablen selektiert. Erwartungsgemäß hat die BS-Bodenart die größte Bedeutung, sie stellt bei den ersten drei Knoten die Trennvariable. Interessanterweise wird die Zustandsstufe nicht zur Schätzung des Tongehaltes herangezogen (sondern zur Schätzung des Schluffgehaltes). Es wäre zu erwarten gewesen, dass

eine Variable, die die Bodenbildung und -entwicklung beschreibt, eine geeignete Variable zur Schätzung des durch Verwitterung und Tonverlagerung beeinflussten Tonanteils darstellt.

Tab. 5.1: Trennvariable und Knotenzahl der verwendeten Regressionsbäume

Zielvariable	Trennvariable	Knotenzahl
Tongehalt	Bodenart, Hangneigungsabschlag, Ackerzahl, Bodenzahl	9
Schluffgehalt	Ackerzahl, Bodenzahl, Zustandstufe, Bodenart, Entstehung, Hangneigungsabschlag	14
Sandgehalt	Ackerzahl, Bodenzahl, Bodenart, Entstehung, Hangneigungsabschlag	14

5.1 Ergebnisse der Schätzungen

Die Ergebnisse der beiden Schätzverfahren sind in Abb. 5.2 für die drei Korngrößenfraktionen Ton, Schluff, Sand in Form von Schätzwert vs. Messwert-Diagrammen dargestellt. Die Abbildungen weisen jeweils akzentuierte Bereiche der einzelnen Korngrößenfraktionsanteile auf, die auf eine bevorzugte Zusammensetzung der Bodenart im Untersuchungsgebiet hinweisen.

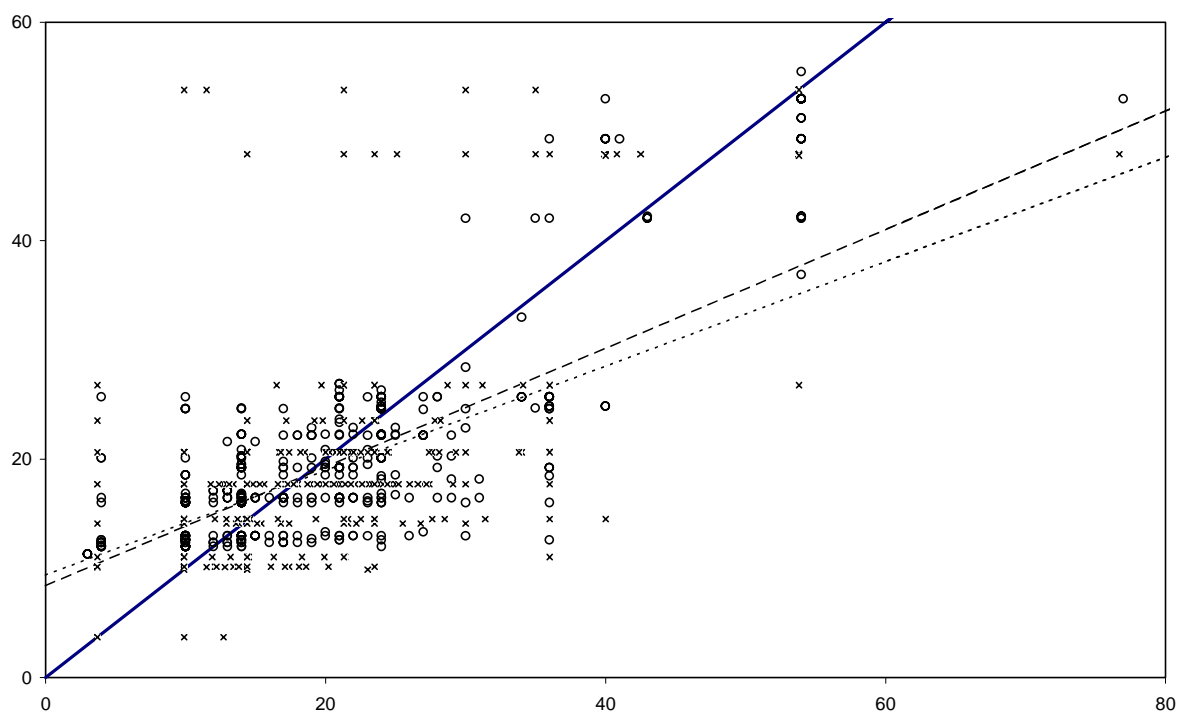


Abb. 5.2: Schätzwerte vs. Messwerte der Verfahren zur Schätzung des Tongehaltes

Erläuterung siehe Abb. 5.3.

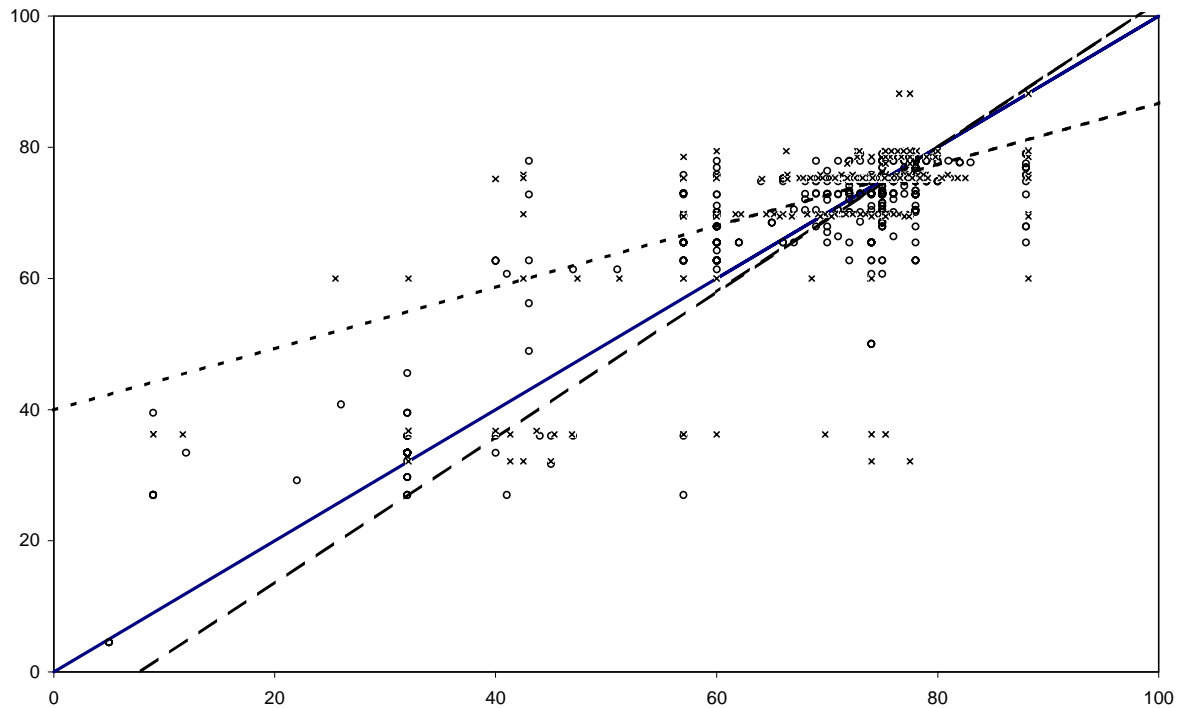


Abb. 5.3: Schätzwerte vs. Messwerte der Verfahren zur Schätzung des Schluffgehaltes

Auf der Ordinate sind die Schätzwerte, auf der Abszisse die Messwerte des Tongehaltes der CART- und Weiblen-Verfahren dargestellt. Eingezeichnet sind die identische Abbildung zur Orientierung und die robusten Ausgleichsgeraden nach Tukey für die beiden Verfahren. Die Parameter der Ausgleichsgeraden sind in Abb. 5.11 vergleichend dargestellt. Abszissen- und Ordinateneinheiten in % (m/m). $N=441$ (CART) bzw. $N=433$ (Verfahren Weiblen). Einige Datenpaare kommen mehrfach vor, daher können mehrere Datenpunkte übereinander liegen. Legende:

° CART Schätzungen

x Verfahren Weiblen

———— 1:1 Beziehung zwischen Schätz- und Messwerten

----- robuste Regressionsgerade nach Tukey der Ergebnisse nach Weiblen

----- robuste Regressionsgerade nach Tukey der CART Ergebnisse

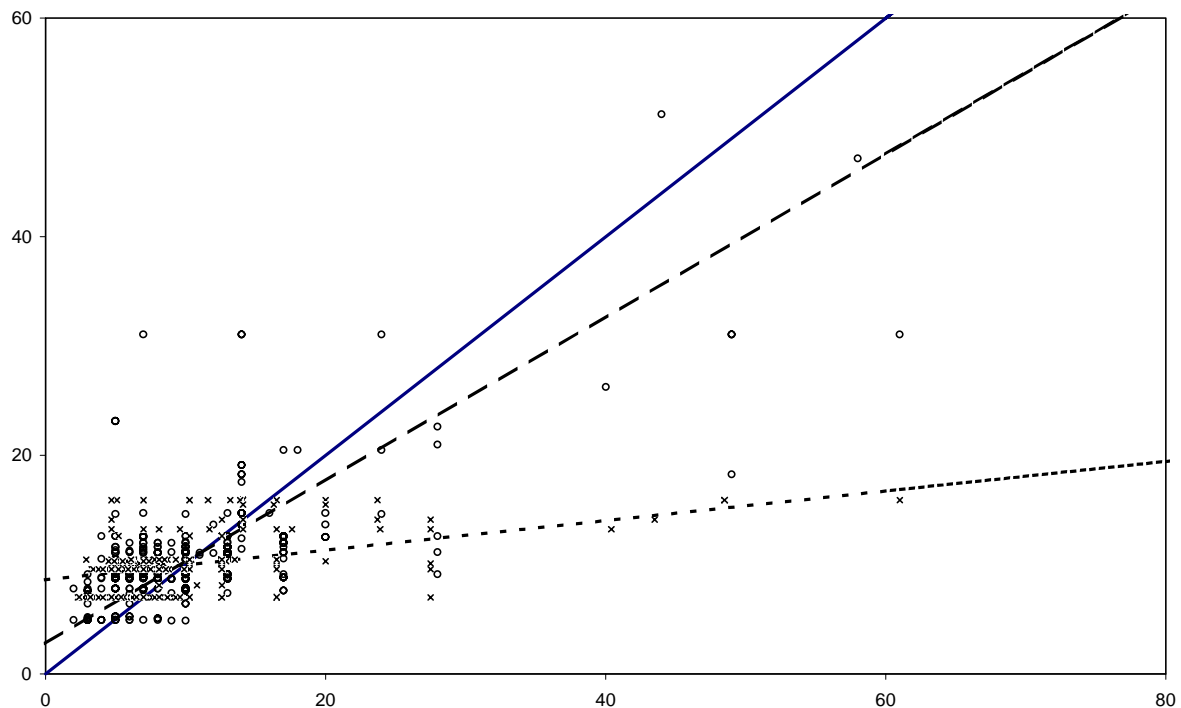


Abb. 5.4: Schätzwerte vs. Messwerte der Verfahren zur Schätzung des Sandgehaltes

Erläuterung siehe Abb. 5.3.

Es existieren teilweise deutliche Unterschiede zwischen den robusten Ausgleichsgeraden und der identischen Abbildung, die eine systematische Über- und Unterschätzung der beiden Schätzverfahren belegen. Zwischen den beiden Schätzverfahren sind Unterschiede erkennbar, die sich insbesondere durch unterschiedliche Ausgleichsgeraden dokumentieren. Bei beiden Schätzverfahren existieren waagerechte und senkrechte Datenstrukturen, die auf gleiche Schätzwerte bei unterschiedlichen Messwerten, bzw. unterschiedliche Schätzwerte bei gleichen Messwerten beruhen und als Indikator für eine zu geringe Differenzierung der Schätzverfahren gelten können.

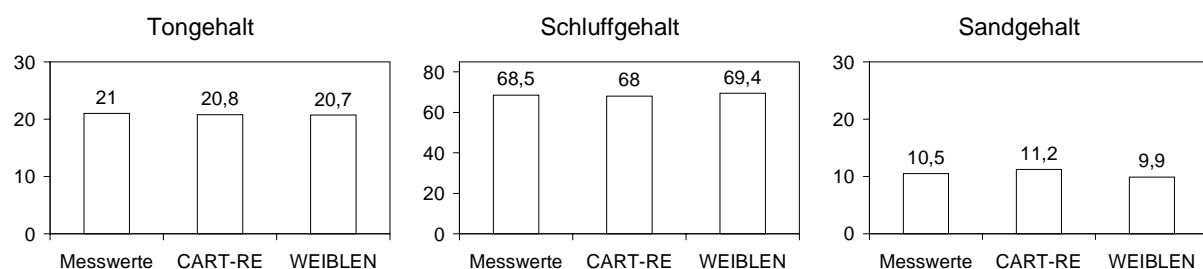


Abb. 5.5: Arithmetische Mittelwerte der Mess- und Schätzwerte der Korngrößenfraktionsanteile

Dargestellt sind die arithmetischen Mittelwerte der Messungen (Datensatz Weiblen, $N=433$) und der Schätzverfahren CART und Weiblen); Ordinateneinheiten in % (m/m).

Die Abweichungen der mittleren Schätzergebnisse der einzelnen Korngrößenfraktionen Ton, Schluff und Sand zu den gemessenen Werten betragen weniger als einen Prozentpunkt. Die mittlere Bodenar-

tenzusammensetzung wird daher durch beide Schätzverfahren gut wieder gegeben. Als – hypothetische – mittlere Bodenart des Untersuchungsgebietes ergibt sich stark toniger Schluff. Aufgrund der Eigenschaften der Verfahren beträgt die Summe der Korngrößenfraktionsanteile 100 Prozent.

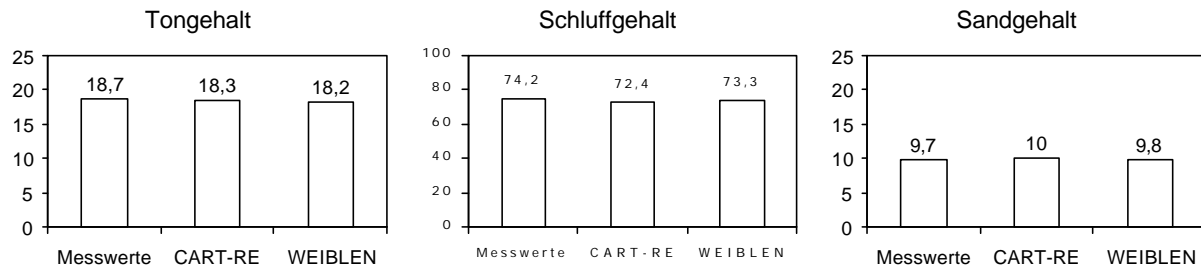


Abb. 5.6: Robuste Mittelwerte der Mess- und Schätzwerte der Korngrößenfraktionsanteile

Dargestellt sind die robusten Mittelwerte der Messungen (Datensatz Weiblen, $N=433$) und der Schätzverfahren CART und Weiblen; Ordinateneinheiten in % (m/m)

Wie beim arithmetischen Mittel gibt es beim robusten Mittelwert ebenfalls geringe Abweichungen zwischen den robusten Mitteln der Mess- und der Schätzwerte. Die Abweichungen betragen unter zwei Prozentpunkte. Die Ergebnisse der Verfahren CART und Weiblen sind beim Ton- und Sandgehalt sehr ähnlich und unterscheiden sich nur beim Schluffgehalt um mehr als 0,1 Prozentpunkte. Bei der Korngrößenfraktion Schluff existiert ein größerer Unterschied zwischen dem arithmetischen und robusten Mittel von über fünf Prozentpunkten, bei den Korngrößenfraktionen Ton und Sand sind die Abweichungen mit maximal 2,3 Prozentpunkten (Tongehalt) deutlich geringer.

Die Summe der einzelnen Korngrößenfraktionsanteile unterscheidet sich im Bereich von wenigen Prozentpunkten. Die Summe der Korngrößenfraktionsanteile der Messwerte beträgt 102,6 %, bei CART 100,7 % und beim Verfahren Weiblen 101,3 %. Das CART Verfahren kommt damit hinsichtlich der robusten Mittelwerte der Korngrößenfraktionsanteile der gewünschten Summeneigenschaft von 100 % mit nur geringem Abstand zum Verfahren Weiblen am nächsten.

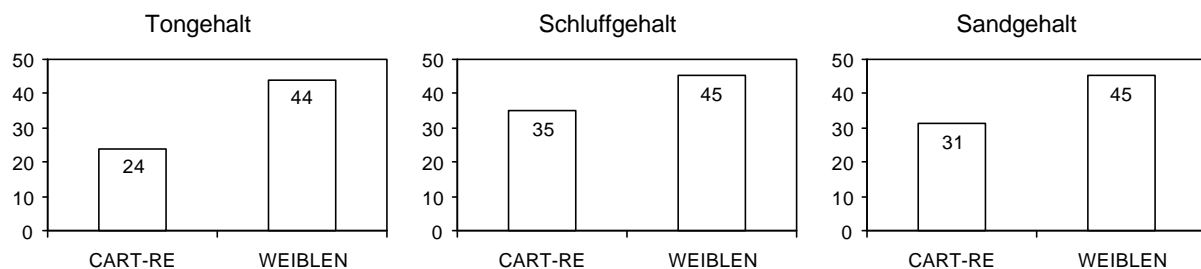


Abb. 5.7: Maximale Schätzabweichung (me) der Schätzungen der Korngrößenfraktionsanteile

Dargestellt sind die maximalen Schätzabweichungen der Schätzverfahren CART und Weiblen; Ordinateneinheiten in % (m/m).

Bei der maximalen Schätzabweichung (me) zeigt sich ein einheitliches Bild: Bei allen Korngrößenfraktionsanteilen ist die maximale Schätzabweichung beim CART-Verfahren deutlich geringer als

beim Verfahren Weiblen. Beim Tongehalt ist der Unterschied am ausgeprägtesten: Die maximale Schätzabweichung ist beim Verfahren Weiblen fast doppelt so groß als beim CART-Verfahren.

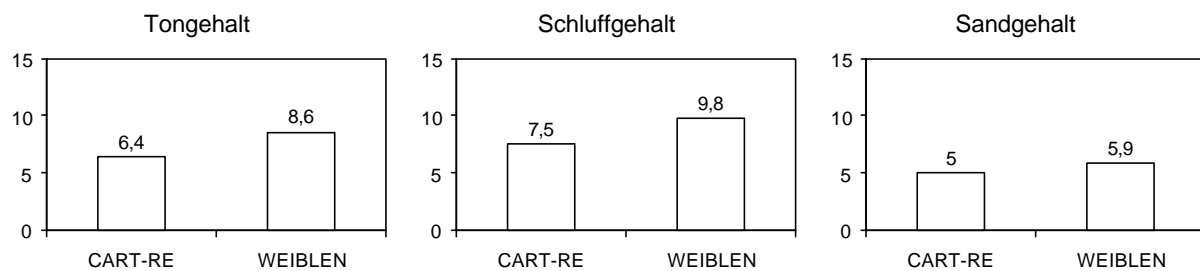


Abb. 5.8: Wurzel der mittleren quadrierten Schätzabweichung ($rmsr$) der Korngrößenfraktionsanteile
 Ordinateneinheiten in % (m/m).

Auch bei der mittleren quadrierten Schätzabweichung ($rmsr$) hat CART immer deutlich geringere Werte als das Verfahren Weiblen. Der deutlichste Unterschied ist auch hier beim Tongehalt festzustellen, beim Sandgehalt ist der Unterschied am geringsten ausgeprägt.

Die durchschnittliche Größe der mittleren quadrierten Schätzabweichung des CART-Verfahrens entspricht etwa der durchschnittlichen Spannweite der Bodenartenuntergruppen des Bodenartendiagramms der Bodenkundlichen Kartieranleitung.

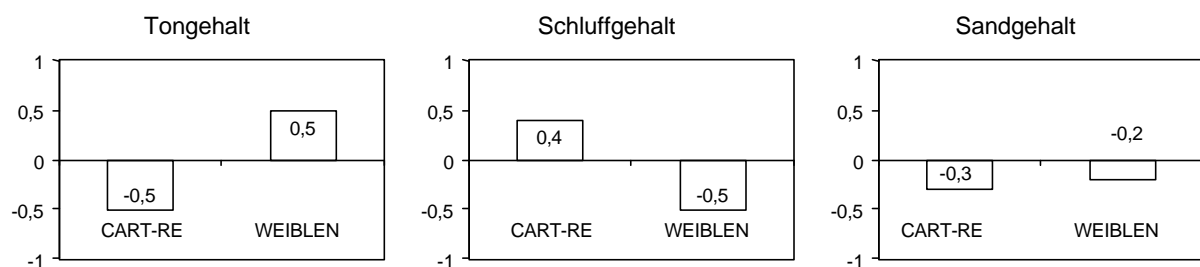


Abb. 5.9: Robustes Mittel der Schätzabweichungen (rer) der Korngrößenfraktionsanteile
 Ordinateneinheiten in % (m/m).

Die robusten Mittelwerte der Schätzabweichungen (rer) der beiden Verfahren sind betragsmäßig annähernd gleich, weisen aber teilweise unterschiedliches Vorzeichen der Abweichung auf. Die Größe des Betrags liegt durchgehend kleiner oder gleich 0,5 Prozentpunkten. Die Tendenz der Schätzabweichungen liegt also nahe bei Null. Interessant ist der Befund annähernd gleicher robuster Schätzabweichungen im Zusammenhang mit den deutlichen Unterschieden bei den maximalen Schätzabweichungen. Offenbar hat das CART-Verfahren seine besondere Stärke im untersuchten Fallbeispiel darin, zu erheblich kleineren maximalen Schätzabweichungen zu führen als das Weiblen-Verfahren.

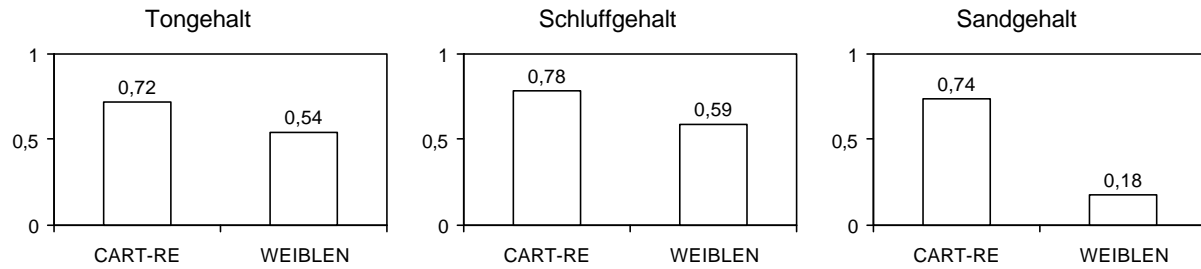


Abb. 5.10: Bestimmtheitsmaß (*cdr*) der Ausgleichsgeraden zwischen Mess- und Schätzwerten

Das Bestimmtheitsmaß (*cdr*) zeigt durchgehend bei allen Korngrößenfraktionen deutlich höhere Werte für CART. Insbesondere beim Sandgehalt ist der Unterschied sehr ausgeprägt.

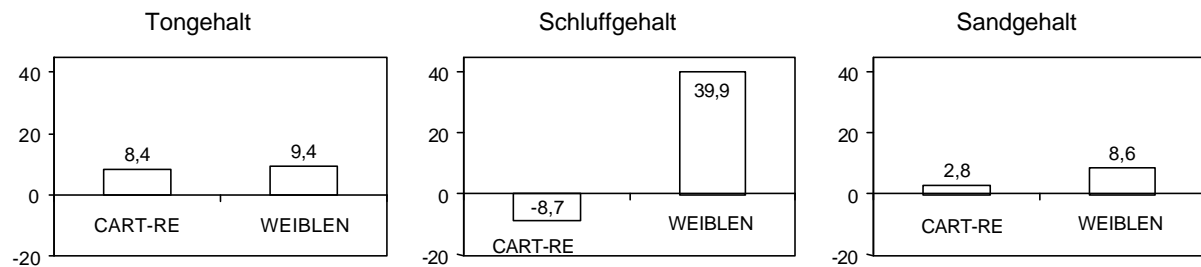


Abb. 5.11: Achsenabschnitt der robusten Regressionsgerade (*a*) zwischen Mess- und Schätzwerten

Regressionsgerade nach Tukey (1977) Ordinateneinheiten in % (m/m).

Beim Achsenabschnitt der robusten Ausgleichsgerade bestehen zwischen den Verfahren teilweise deutliche Unterschiede. Die größten Unterschiede liegen beim Schluff- und beim Sandgehalt vor. Der Schluffgehalt wird von CART bei niedrigen Werten unterschätzt, vom Weiblen-Verfahren hingegen massiv überschätzt. Beim CART-Verfahren liegt ein vergleichsweise geringer Achsenabschnitt bei Sandgehalt vor. Das CART-Verfahren hat bei allen drei Korngrößenfraktionen einen betragsmäßig kleineren Wert als das Weiblen Verfahren.

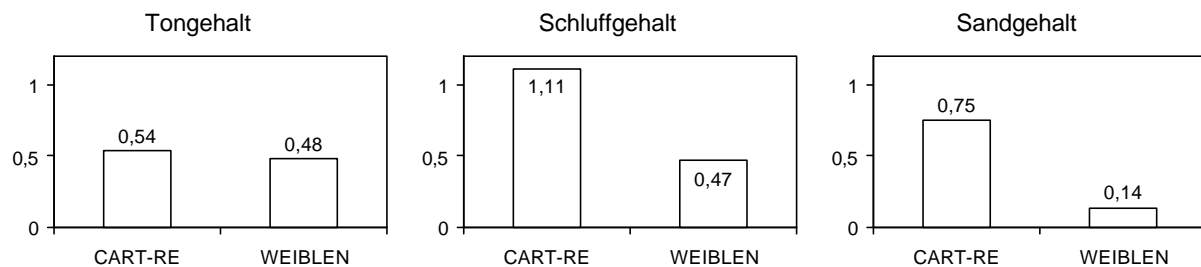


Abb. 5.12: Steigung der robusten Regressionsgerade (*b*) zwischen Mess- und Schätzwerten

Regressionsgerade nach Tukey (1977)

Auch die Steigung der robusten Ausgleichsgerade ist durch sehr große Unterschiede zwischen den Verfahren bei Schluff- und Sandgehalt gekennzeichnet. Beim Tongehalt sind die Unterschiede weni-

ger ausgeprägt. Das CART-Verfahren hat bei allen drei Korngrößenfraktionen einen deutlich näher am Optimalwert Eins liegenden Wert als das Weiblen Verfahren. Die Steigungen schwanken dabei beim CART-Verfahren von 0,5 bis 1,1.

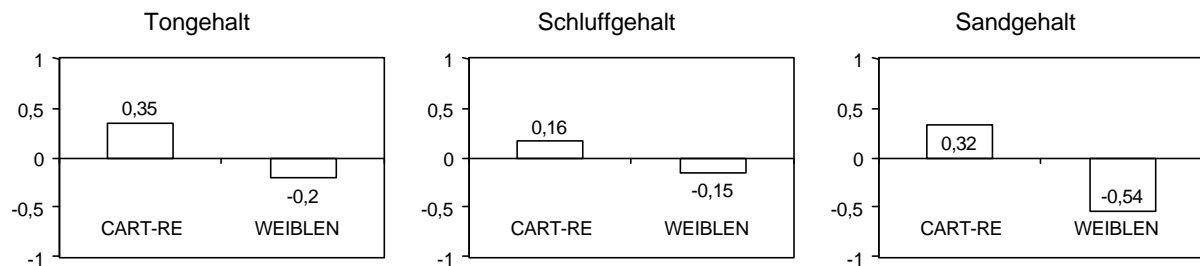


Abb. 5.13: Koeffizient der maximalen Schätzabweichung (*comare*)

Die Koeffizienten der maximalen Schätzabweichung (*comare*) sind bei CART alle positiv, die des Verfahrens Weiblen sind teilweise stark negativ. Daher scheint CART hinsichtlich der Minimierung der maximalen Schätzabweichung deutlich besser geeignet zu sein. Am stärksten ausgeprägt sind die Unterschiede zwischen beiden Verfahren beim Sandgehalt, wo das Weiblen Verfahren eine um über 50 % größere maximale Schätzabweichung hat als bei Verwendung des arithmetischen Mittels aus größtem und kleinstem Messwert als konstante Schätzung.

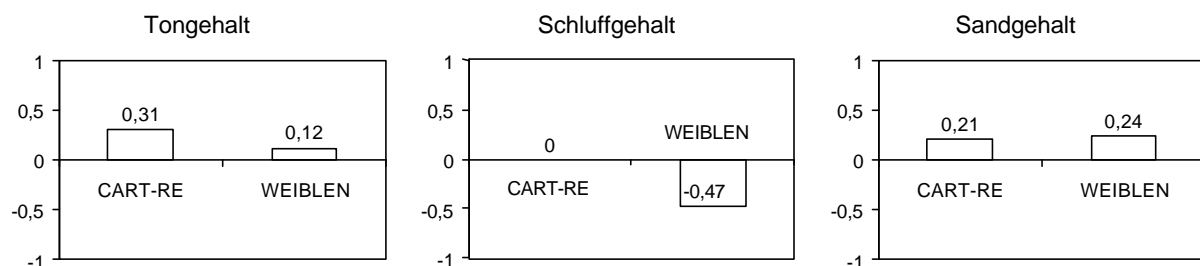


Abb. 5.14: Koeffizient der zentralen Schätzabweichung (*cocare*)

Auch bei der zentralen Schätzabweichung (*cocare*) hat CART hinsichtlich des Ton- und Schluffgehaltes größere Koeffizienten aufzuweisen als das Verfahren Weiblen. Einzig beim Sandgehalt hat das Weiblen-Verfahren einen leicht größeren Wert. Beim Schluffgehalt findet sich als größter Koeffizient der beiden Verfahren bei CART der Wert Null, d. h. dass bei Verwendung des Medians als Schätzverfahren mit konstanter Schätzung die gleiche zentrale Schätzabweichung resultiert wie beim CART-Verfahren. Insofern kann der Median der Messwerte selbst als ein geeignetes Schätzverfahren bezüglich des Schluffgehaltes gelten.

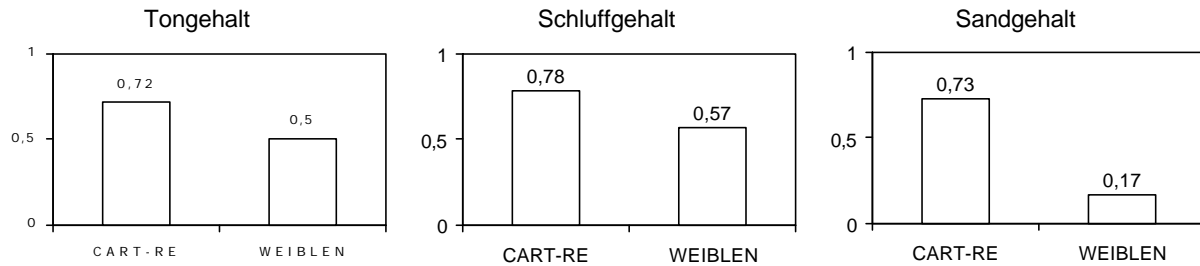


Abb. 5.15: Koeffizient der mittleren quadrierten Schätzabweichung (*comsre*)

Die Koeffizienten der mittleren quadrierten Schätzabweichung (*comsre*) weisen unter allen *core*-Koeffizienten mit Abstand die größten Werte auf. Die Reduktion der mittleren quadrierten Schätzabweichung gegenüber dem Mittelwert der Messungen als Schätzverfahren beträgt bei CART mindestens 72 %. Auch bei *comsre* erzielt das CART-Verfahren bei allen Korngrößen größere Koeffizienten als das Weiblen-Verfahren.

Die großen *comsre*-Werte zeigen, dass das arithmetische Mittel der Messwerte als Schätzverfahren wesentlich größere quadrierte Schätzabweichungen aufweist als das CART-Verfahren.

Erwartungsgemäß liegen die Koeffizienten der kreuzvalidierten Schätzung niedriger als die der Resubstitutionsschätzung, sie sind jedoch bei allen Korngrößenfraktionen deutlich größer als die des Weiblen Verfahrens. Es besteht jedoch ein deutlicher Unterschied zwischen der Resubstitutions- und der Kreuzvalidierungsvariante, der trotz der großen Stichprobenanzahl von über 400 Datenelementen auf instabile Regressionen schließen lässt.

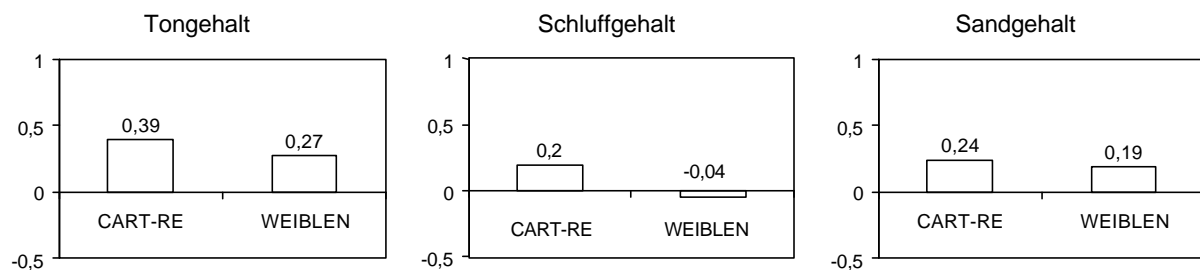


Abb. 5.16: Koeffizient der robusten Schätzabweichung (*corore*)

Beim Koeffizienten der robusten Schätzabweichung (*corore*) ist das Niveau der Werte geringer als bei *comsre*. Das CART-Verfahren hat bei allen drei Korngrößenfraktionen positive Werte, die – wie bei den drei vorangegangenen *core*-Kriterien – größer sind als die des Weiblen-Verfahrens.

5.2 Zusammenfassung der Ergebnisse

Die drei Korngrößenfraktionen unterscheiden sich hinsichtlich der Gütekriterien voneinander. Zwischen CART und dem Weiblen Verfahren gibt es größere Unterschiede. CART hat bei allen drei Korngrößenfraktionen einen stärker ausgeprägten linearen Zusammenhang (Bestimmtheitsmaß, Steigung, Achsenabschnitt) zwischen den gemessenen und den geschätzten Werten als das Weiblen-Verfahren. Bei allen dargestellten Kriterien zeigt sich, dass das CART-Verfahren dem Weiblen-

Verfahren überlegen ist. Am deutlichsten sind die Unterschiede zwischen den beiden Verfahren bei der maximalen Schätzabweichung, wo sich beim Weiblen-Verfahren fast doppelt so große Abweichungen finden wie bei CART. Die Unterschiede der beiden Verfahren zwischen dem robusten Mittel der Schätzabweichungen und der maximalen Schätzabweichung deuten darauf hin, dass die Stärken des CART Verfahrens insbesondere in geringeren maximalen Schätzabweichungen – verglichen mit dem Weiblen-Verfahren – zu finden sind.

Die *core*-Koeffizienten unterscheiden sich teilweise deutlich voneinander. Beim Schluffgehalt beispielsweise beträgt *comsre* beim Weiblen-Verfahren 0,57, *cocare* hingegen $-0,47$. Negative *core*-Werte kommen beim Weiblen-Verfahren vor, was bedeutet, dass einfache konstante Schätzwerte als Schätzverfahren eine geringere Schätzabweichung aufweisen als das Weiblen Verfahren.

5 Diskussion

5.1 Vergleich von CART mit einfachen Referenzverfahren

Die maximale Schätzabweichung der CART Schätzungen ist bei den beiden Zielvariablen Nitratgehalt im Boden und Korngrößenfraktionsanteile geringer als die der meisten Vergleichsverfahren. Dies trifft in besonderem Maße für den Vergleich der kreuzvalidierten Varianten zu. CART nimmt daher hinsichtlich der Minimierung der maximalen Schätzabweichung im Verfahrensvergleich aller untersuchter Verfahren eine Position im oberen Bereich ein.

Sowohl das arithmetische als auch das robuste Mittel werden von CART in beiden Fallstudien etwa ähnlich gut (Schätzung der Korngrößenfraktionsanteile) oder besser (Nitratgehalt-Werte) wiedergegeben als bei den Vergleichsverfahren. CART zeigt somit in Hinsicht auf Erwartungstreue und robuste Tendenz bei beiden Fallstudien positive Eigenschaften.

Das Verhältnis zwischen gemessenen und geschätzten Werten ist bei beiden Fallbeispielen durch größere Steigungen und kleinere Achsenabschnitte der Ausgleichsgeraden der CART-Werte gegenüber den Ausgleichsgeraden der konkurrierenden Verfahren gekennzeichnet.

Die *core*-Koeffizienten sind beim CART-Verfahren in beiden Fallbeispielen fast ausnahmslos positiv (Ausnahme: *corore*), d. h. die normierten Schätzabweichungen sind fast immer kleiner als bei Verwendung eines konstanten Wertes als Schätzverfahren. Die Vergleichsverfahren hingegen weisen meistens deutlich größere normierte Schätzabweichungen auf, die manchmal sogar größer sind als bei Verwendung eines konstanten Wertes.

Die Kombination von geringerer Verzerrung, geringerer maximaler Schätzabweichung und allgemein besserem Verhältnis der geschätzten vs. der gemessenen Werte (Steigung, Achsenabschnitt der Ausgleichsgeraden, Summeneigenschaft der Korngrößenfraktionsanteile) und quasi immer geringerer Schätzabweichung als bei Verwendung von geeigneten konstanten Werten bei den CART-Regressionsbäumen deutet insgesamt auf eine bessere Nutzung der zu Verfügung stehenden Informationen und größere Plausibilität der CART-Regressionsbäume hin. Daher erscheinen die CART-Regressionsbäume bei den vorliegenden Fallbeispielen die bessere Eignung als die untersuchten Vergleichsverfahren zu besitzen.

Verglichen mit sehr elaborierten Kriging Verfahren, die zur Schätzung der Bodenfeuchte im Untersuchungsgebiet von Lehmann (1995) verwendet werden konnten, bewegen sich die bei der Variable Nitratgehalt erzielten *comsre*-Werte der kreuzvalidierten CART-Variante auf gleichem, die bei der Variablen Korngrößenfraktionsanteile auf deutlich höherem Niveau. In Anhang 5 ist eine Zeitreihe der Schätzung der Bodenfeuchte im gleichen Untersuchungsgebiet mit dem Bayes-Markow-Kriging-Verfahren dargestellt. Als zentraler Wert für *comsre* ergibt sich für die Zeitreihe vom 14.2.1990 – 13.1.1991 ein Wert von 0,12 (arithmetisches Mittel ebenfalls 0,12). Verglichen mit dem erheblichen Aufwand für die Beprobung (45.000 Einzelmessungen), die Variogrammschätzung und das sehr komplexe Kriging-Verfahren unter Verwendung von Zusatzinformationen ist das CART-Verfahren weniger aufwändig und leichter nachvollziehbar. Ob die Möglichkeit, Schätzfehler räumlich verteilt darzustellen, nur auf die Kriging Verfahren beschränkt ist, muss erst noch genauer untersucht werden, bislang wurde hierzu noch kein Versuch unternommen. Denkbar wäre jedoch, bei Zuordnung der

Endknoten des Regressionsbaumes zu den Flächenelementen die Schätzung des Schätzfehlers durch die Varianz dieser Endknoten vorzunehmen.

Grundsätzlich sind die Werte der Kreuzvalidierungsvarianten von CART kleiner als die der Resubstitutionsvarianten. Auch bei der Schätzung der Korngrößenfraktionsanteile bestehen trotz der vergleichsweise großen Stichprobenanzahl (über 440 vs. 48 bei Nitratgehalt) deutliche Unterschiede zwischen den beiden Varianten. Offensichtlich sind die Regressionsbäume aufgrund eines zu geringen Stichprobenumfanges instabil gegenüber einer Änderung des Datenensembles.

Die Ergebnisse der beiden Fallstudien deuten darauf hin, dass die CART-Regressionsbäume die Eigenschaft besitzen, zu deutlich geringeren maximalen Schätzabweichungen zu führen als die untersuchten einfachen Vergleichsverfahren. Diese Eigenschaft machen sie für spezielle Fragestellungen mit dieser Zielrichtung interessant und wertvoll.

5.2 Gütekriterien

Die *core*-Koeffizienten führen zu Gütemaßen, die aufgrund ihrer Normierung durch das jeweilige Abstandsmaß nicht mehr von der Variabilität der Messwerte abhängen.

Die Fallbeispiele haben gezeigt, dass mit Hilfe der Gütekriterien ein Vergleich verschiedener Schätzverfahren in Hinsicht auf unterschiedliche Fragestellungen möglich ist.

Die verwendeten Kriterien sind praktikabel und leicht verständlich, sie erfordern nur geringen Rechenaufwand, der sich mit gewöhnlichen Tabellenkalkulationsprogrammen durchführen oder in klassischen Programmiersprachen leicht programmieren lässt.

Die teilweise erheblichen Unterschiede (z. T. mit unterschiedlichem Vorzeichen) zwischen verschiedenen *core*-Kriterien beim selben Verfahren belegen die Notwendigkeit, zur Charakterisierung eines Schätzverfahrens die verschiedenen Maße der Schätzabweichungen zu untersuchen. Darüber hinaus ermöglichen die verschiedenen Gütemaße die Festlegung von Qualitätsstandards für Schätzungen, z. B. die Eigenschaft, dass alle *core*-Kriterien über einem bestimmten Schwellenwert (z. B. Null) liegen sollten.

Für eine fachlich fundierte Auswahl der Schätzverfahren ist es unbedingt erforderlich, die spezifischen Gütekriterien festzulegen, mit denen Schätzungen beurteilt werden sollen. Die Festlegung der Gütekriterien sollte dabei a priori erfolgen, um subjektive Entscheidungsfindungen einzuschränken.

Die *cocare* und *corore*-Werte haben in beiden Fallstudien ein geringeres Niveau als die *comare*- und *comsre*-Koeffizienten. Die *cocare*-Koeffizienten haben im Allgemeinen das kleinste Niveau. Insofern erscheint *cocare* als besonders empfindliches Kriterium geeignet zur Definition von qualitativen Minimalstandards für Schätzverfahren.

Zur Beurteilung der Ergebnisse der räumlichen Schätzverfahren steht somit ein Set von verfahrensinernen und selbstgewählten Gütemaßen zur Verfügung, die damit eine Forderung von Nestler (1999) erfüllen.

5.3 Kreuzvalidierung als Verfahren zur räumlichen Güteschätzung

Die vorliegende Untersuchung unterstützt die These, dass die Kreuzvalidierung ein unabdingbares Instrument darstellt, um die Güte der räumlichen Schätzung abzuschätzen; der Befund von Lehmann (1995) kann insofern bestätigt werden. Selbst bei vergleichsweise (im Bereich der experimentellen Geowissenschaften) großen Datensätzen mit Stichprobenumfängen von über 400 (bei Schätzung der

Korngrößenfraktionsanteile) gibt es jedoch merkliche Unterschiede zwischen der Resubstitutions- und der Kreuzvalidierungsschätzung. Die Unterschiede sind dabei im wesentlichen auf relativ seltene, extreme Werte zurückzuführen, die entsprechend schwach besetzte Endknoten mit großer Varianz zur Folge haben. In diesem Falle dürfte auch die Verwendung von Bootstrap-Verfahren zur Abschätzung der Güte nur zu wenig anderen Ergebnissen führen.

5.4 Robuste Methoden

Robuste Methoden sind hilfreich zur Aufdeckung von Datentendenzen, was bei fehlerbehafteten Daten bei Anwendung nichtrobuster deskriptiver Statistik nur sehr aufwändig zu realisieren ist.

Mit Hilfe des robusten Mittelwertes ist z. B. eine differenziertere Sichtweise der Nitrat-Verteilung im Bodenprofil möglich als bei alleiniger Betrachtung des arithmetischen Mittels. Bei Betrachtung des arithmetischen Mittels (Abb. 4.5) ergibt sich eine Zunahme der Nitratgehalte mit zunehmender Bodentiefe, was aufgrund des robusten Mittels nicht bestätigt werden kann (Abb. 4.6).

Die Analyse der Schätzabweichungen mit robusten und nichtrobusten Verfahren (*rer*, maximale Schätzabweichung) beim zweiten Fallbeispiel charakterisieren die besondere Eigenschaft der CART-Regressionsbäume, zu vergleichsweise geringen maximalen Schätzabweichungen zu führen.

Das robuste Gütekriterium *corore* führt bei beiden Fallbeispielen zu gleichlautenden Ergebnissen wie die nichtrobusten *core*-Kriterien.

Robuste Statistiken (z. B. Huber-M-Schätzer zur Schätzung des robusten Mittelwertes) erscheinen besonders geeignet als Eingangswerte für Anwendungen der Fuzzy-Set-Theorie (Zadeh, 1965; McBratney & Odeh, 1997), da sie repräsentative Werte darstellen, die das Datenensemble gut wiedergeben ohne von einzelnen Extremwerten dominiert werden zu können. Sie sind daher insbesondere für Schätzungen des α -cut 1,0 prädestiniert.

Die Beurteilung der robusten Verfahren ist bislang problematisch, da deren Eigenschaften in Zusammenhang mit Gütekriterien nur unzureichend untersucht sind. Eine detailliertere theoretische Arbeit hierzu hätte den Rahmen einer geowissenschaftlichen Arbeit sowohl inhaltlich als thematisch gesprengt und bleibt daher späteren Arbeiten vorbehalten.

5.5 N_{\min} -Beprobungsmethode

Schmidthaler et al. (1991) finden auf einem nach Silomais brach liegenden Ackerstandort einen analysebedingten Variationskoeffizienten von 5 %. Dieser ist im Verhältnis zum räumlichen Variationskoeffizienten von 30-47 % klein. Die Abweichungen aufgrund räumlich unterschiedlicher Lagerungsdichte und Skelettgehaltes, die in die Berechnung des flächenbezogenen mineralischen Stickstoffgehaltes eingehen und auf Grund der Fehlerfortpflanzung zu Ungenauigkeiten der Bestimmung führen, werden von ihnen auf bis zu 26 % angegeben. Ilseman et al. (2001) ermitteln auf drei untersuchten Standorten in Niedersachsen bei Probenahme gemäß LUFA-Vorschrift (Hoffmann, 1991) ebenfalls Variationskoeffizienten von 31-36 %. Thres et al. (1998) finden auf zwei Standorten in Thüringen bei Annahme einer Lognormalverteilung der Nitratgehalte einen auf die untersuchte Schicht bezogenen Variationskoeffizienten von 12-42 %. Mit der Mineralisation im Frühjahr steigt die auf das Gesamtprofil bezogene Variabilität ihrer Gesamtprofile von 29 % im Herbst auf 67 % im Frühjahr. Nach einer Frühjahrsdüngung nimmt die schichtenbezogene Variabilität zu, nach Umbruch im Winter und dem dadurch bedingtem Mineralisierungsschub nimmt sie hingegen ab. Sie schätzen

den Variabilitätsfehler bei einem Stichprobenumfang von 16, einem mittlerem Stickstoffgehalt von 50 kg N/(ha x 0.3 m) und bei einer Irrtumswahrscheinlichkeit von $\alpha = 0.05$ bei Variationskoeffizienten von 15 bzw. 45 % auf 14 bzw. 50 kg N/(ha x 0.3 m).

Mit Rücksicht auf die Probenahme im Februar vor der Frühjahrsdüngung und der Hauptmineralisationsphase bei allerdings geringerem Stichprobenumfang von 9 bei der vorliegenden Untersuchung, wird aufgrund der Befunde von Thres et al. (1998) vermutet, dass die Ungenauigkeit größer ist als 14 kg N/(ha x 0.3 m). Die Wurzel der quadrierten Schätzfehler (*rmsr*) der Nitratgehalts-Schätzungen von CART bewegt sich im Rahmen der Genauigkeit, mit der die Messungen durchgeführt wurden. Eine weitere Verbesserung der Schätzungen erscheint auf Grundlage der vorhandenen Daten bzw. der praktizierten Probenahmetechnik daher nicht möglich. Eine umfassende Beurteilung der Schätzverfahren, über die Darstellung und Diskussion der Gütekriterien hinaus, ist daher nicht möglich.

Es ist unklar, ob bei genauer bestimmten Werten eine wesentliche Verbesserung der Schätzungen möglich ist.

In Hinsicht auf eine praktische Anwendung in der Landwirtschaft sind daher andere Methoden zur Bestimmung der Stickstoffversorgungssituation von Pflanzen erforderlich, wie sie von verschiedenen Autoren (Daughtry, et al., 2000; Delgado et al., 2001) erprobt werden (z. B. Chlorophyll-Bestimmung, Pflanzensaftanalyse mit Nitrat-Elektroden).

6 Schlussfolgerungen

Die von Lehmann (1995) formulierten Anforderungen an räumliche Schätzverfahren für die Bestimmung der Bodenfeuchte in einem kleinen ländlichen Einzugsgebiet werden in Hinsicht folgender Aspekte erfüllt:

- Hohe räumliche Auflösung der Schätzungen
- Objektive Beurteilungskriterien
- Vertretbarer zeitlicher Aufwand

Die räumliche Auflösung der CART-Schätzungen ist im wesentlichen nur von der Auflösung der Zusatzinformationen abhängig. Insofern ist bei Verwendung von entsprechenden Zusatzvariablen eine gleiche räumliche Auflösung wie bei dem von Lehmann (1995) praktizierten Bayes-Markov-Kriging realisierbar.

Als Entscheidungsgrundlage für die Gütebeurteilung wurden im Rahmen dieser Untersuchung eine Reihe neuer Gütekriterien eingeführt, die eine klarere und differenziertere Charakterisierung ermöglicht als die von Lehmann verwendeten Gütekriterien. Aufgrund der Normierung der *core*-Kriterien können objektive Verfahrensvergleiche durchgeführt und Minimalstandards für die Schätzungen festgelegt werden.

Die Anwendung des CART-Verfahrens ist weniger zeitintensiv als die des Bayes-Markov-Krigings. Das Kriging-Verfahren erfordert eine sehr zeitaufwändige Variogrammbestimmung, die bei CART-Regressionsbäumen nicht notwendig ist. Beide Verfahren setzen jedoch Expertenwissen und Erfahrung mit dem Schätzverfahren voraus.

Die Ergebnisse von Lehmann (1995) hinsichtlich der Güte der Nächste-Nachbar-Schätzung können insofern bestätigt werden, als die Nächste-Nachbar-Schätzung bei den meisten Gütekriterien am schlechtesten abschneidet.

Vergleichbar mit dem von Mertens (2000) beschriebenen Verfahren der Fehlerschätzung bei Anwendung der Fuzzy-Set-Theorie zur Schätzung des Nitratgehaltes der Sickerwassers können bei CART-Regressionsbäumen beliebige Variabilitätsmaße der Endknoten berechnet werden und zur Fehlerschätzung herangezogen werden.

Weiche Informationen, bzw. Expertenwissen (Mertens, 2000) können in einem Black-Box-Regressionsmodell, wie die CART-Regressionsbäume, nur in Form einer nominalskalierten Variable integriert werden. Eindeutige Lösungen im Sinne eines von der Wurzel an weitgehend gleichartigen Baumaufbaus sind ebenfalls nicht zwingend, bei Bäumen mit unterschiedlichem Aufbau sind durchaus ähnliche Schätzergebnisse möglich.

Der methodische Ansatz von CART ist intuitiv leicht zu erschließen und zu vermitteln. Da der Baumaufbau jedoch nicht unbedingt kausal nachvollziehbar ist, können sich Glaubwürdigkeitsprobleme ergeben.

Die Normierung der *core*-Kriterien ist für die Verständlichkeit der Gütekriterien und die Bewertung der Schätzverfahren vorteilhaft. Die Variabilitätsmaße der Werte der Endknotenelemente als Maß für die Fehlerschätzung lassen sich im Gegensatz zu den *core*-Koeffizienten flächenhaft in Kartenform darstellen.

Die Übertragbarkeit der CART-Methode auf andere Untersuchungsgebiete ist gegeben, jedoch ist der Vorgang der Baumerstellung und –optimierung in jedem anders gearteten Untersuchungsgebiet erneut erforderlich.

CART ist ein parameterfreies Verfahren. Daher sind die Klassifikations- und Regressionsbäume im Gegensatz zu parametrischen Verfahren wenigstens prinzipiell geeignet, falls fehlerbehaftete Datensätze vorliegen. Allerdings ist die Effizienz (Pitman, 1949) eines parameterfreien Verfahrens geringer als die eines robusten Verfahrens.

Zwischen den beiden untersuchten Zielvariablen Nitratgehalt und Korngrößenfraktionsanteile gibt es hinsichtlich der Stichprobenanzahl und der Gütemaße große Unterschiede. Ein Zusammenhang zwischen Stichprobenanzahl und Güte ist allerdings aufgrund der wenig repräsentativen Nitratwerte nicht nachweisbar. Trotz der etwa zehnfachen Stichprobenanzahl im Vergleich zur Schätzung der Nitratwerte existiert auch bei der Schätzung der Bodenart ein deutlicher Unterschied zwischen der Kreuzvalidierungs- und der Resubstitutionsschätzung der Schätzabweichungen, d. h. die Regressionsbäume der Resubstitutions- oder Kreuzvalidierungsschätzung unterscheiden sich deutlich voneinander. Die vorliegende Untersuchung unterstützt daher die Forderung von Nestler (1999) nach weiterer Forschung hinsichtlich der Stabilität des Entscheidungsbaumes in Abhängigkeit des Stichprobenumfanges.

Ebenso können die Ergebnisse von Nestler (1999) insofern bestätigt werden, als auch die Qualität von Schätzungen mit Hilfe der in dieser Schrift verwendeten Regressionsbäumen im Wesentlichen durch die Unschärfe der Schätzvariablen limitiert wird.

Aufgrund der vorliegenden Ergebnisse wird eine andere Sichtweise als die von Kraft (1998) und Nestler (1999) vertreten, was die Vergleichbarkeit der Gütemaße betrifft. Es existieren markante Unterschiede zwischen den von Kraft und Nestler verwendeten Klassifikationsbäumen und den in dieser Arbeit angewandten Regressionsbäumen, z. B. was die Möglichkeit der Voreinstellung von a priori Wahrscheinlichkeiten angeht. Die Kreuzvalidierungskosten sind durchaus mit den Resubstitutionskosten vergleichbar, der Unterschied besteht lediglich in einer sich gering unterscheidenden Trainingsmenge. Aus etwaigen Unterschieden zwischen Resubstitutions- und Kreuzvalidierungskosten können

sehr wohl Aussagen über die Stabilität des Regressionbaumes gezogen werden. Sind diese Unterschiede sehr klein, ist auch der Regressionsbaum recht stabil. Bei einem größerem Unterschied, wie er z. B. bei den beiden hier untersuchten Fallbeispielen vorliegt, unterscheiden sich die betreffenden Regressionsbäume deutlich voneinander. Der Befund, dass hinsichtlich der Kreuzvalidierungskosten ähnlich bewertete Bäume unterschiedliche Karten produzieren, ist nicht als methodischer Nachteil der internen CART-Gütekriterien zu verstehen, sondern die Folge einer quantitativ nicht ausreichenden Datengrundlage. In diesem Falle ermöglicht der Datenraum mehrere Klassifikations- oder Regressionsbäume mit ähnlichen Kreuzvalidierungskosten. Insofern sind mangelhafte, bzw. nicht plausibel erscheinende Ergebnisse auf die Größe und Heterogenität der Datenbasis zurückzuführen (Kraft, 1998; Nestler, 1999).

Die vorliegenden Fallbeispiele führen zur Schlussfolgerung, dass die Datenbasis für flächenhafte Schätzungen sowohl qualitativ (insbesondere bei Nitratgehalt) als auch quantitativ als sehr gering zu bewerten ist. Dieses Ergebnis steht in Übereinstimmung mit den Befunden von Nestler (1999) und Kraft (1998). Zumindest in quantitativer Hinsicht sind hierzu auch die Ergebnisse von Lehmann (1995) zu zählen, der mit sehr komplexen und aufwändigen Verfahren nur eine mittlere Varianzreduktion im Bereich von 10-15 % erzielen konnte.

Die Qualität der Schätzungen ist untrennbar mit einer qualitativ und quantitativ ausreichenden Datengrundlage verbunden. Eindeutigkeitsprobleme und schlechte Schätzleistungen sind bei komplexen Zielvariablen, die in den Geowissenschaften untersucht werden, insbesondere auf mangelhafte Datengrundlage zurückzuführen. Mit Regionalisierungsverfahren können keine Wunder gezaubert werden, sondern sie erfordern eine weit bessere Datengrundlage als in den Geowissenschaften bislang allgemein akzeptiert. Es zeigt sich, dass bei vielen Regionalisierungsansätzen, so z. B. im Weiherbachprojekt, mit deutlich unter 100 Datenelementen allenfalls eine räumliche Schätzung erzielt werden kann, die substantiell kaum besser ist als bei Verwendung eines konstanten Wertes z. B. des Mittelwertes, des Medians oder des robusten Mittelwertes (die als Referenzschätzer bei den *core*-Kriterien implementiert sind).

Die Unterschätzung des Regionalisierungsaufwandes macht es notwendig, die Strategie, flächenhaft regionalisierte Parameter als Grundlage für deterministische Modellierungen zu verwenden, neu zu überdenken. Der Aufwand für die Bereitstellung der flächenhaft regionalisierten Parameter wurde in der Vergangenheit offenbar unterschätzt. Insofern ist eine Neubewertung der flächenhaften deterministischen Modellierungsansätze erforderlich.

Auf Grund der dargestellten Befunde erscheint eine Regionalisierung von Bodenstickstoffgehalte, die nach der N_{\min} Methoden beprobt wurden, mit vertretbarem Aufwand nicht möglich.

Die N_{\min} -Beprobungsmethode stellt sich für wissenschaftliche Zwecke als nicht praktikabel dar, da bei der klassischen Beprobung extrem hohe Stichprobenumfänge für hinreichend genaue Ergebnisse erforderlich sind. Auch für praktische Anwendungen in der Landwirtschaft sind die Ergebnisse als Ausgangsbasis für eine bedarfsgerechte Düngung sehr ungenau.

Literaturverzeichnis

- Addiscott, T.M., Wagenet, R. J. 1985. Concepts of solute leaching in soils: A review of modelling approaches. *J. Soil Sci.* 36: 411-424.
- Addiscott, T.M., Whitmore, A. P. 1987. Computer simulation of changes in soil mineral nitrogen and crop nitrogen during autumn, winter and spring. *J. Agric. Sci.* 109: 141-157.
- Agterberg, F. P. 1984. Trend surface analysis. In Gaile G. L., Willmott, C. J. (ed.) *Spatial statistics and models* Reidel, Dordrecht, Niederlande, 14-171.
- Allewell, C., Manderscheid, B. 1998. Use of objektive criteria for the assessment of biogeochemical ecosystem models. *Ecological modelling* 107: 213-224.
- Andrews, D.F., Bickel, P.J., Hampel, F.R., Huber, P.J., Rogers, W.H., Tukey J.W. 1972. *Robust Estimates of Location: Survey and Advances*. Princeton University Press, Princeton N.Y.
- Arbeitsgruppe Bodenkunde 1982. *Bodenkundliche Kartieranleitung*. Hrsg: Bundesanstalt für Geowissenschaften und Rohstoffe, Geologische Landesämter in der Bundesrepublik Deutschland 3. Auflage. Schweizerbart'sche Verlagsbuchhandlung Stuttgart Hannover.
- Arbeitsgruppe Bodenkunde 1996. *Bodenkundliche Kartieranleitung*. Hrsg: Bundesanstalt für Geowissenschaften und Rohstoffe, Geologische Landesämter in der Bundesrepublik Deutschland. 4. Auflage. Schweizerbart'sche Verlagsbuchhandlung Stuttgart Hannover.
- Arens, H. 1960. Die Bodenkarte 1:5.000 auf der Grundlage der Bodenschätzung, ihre Herstellung und ihre Verwendungsmöglichkeiten. Reihe: Fortschritte in der Geologie von Rheinland und Westfalen, Band 8. Geologisches Landesamt Nordrhein-Westfalen, Krefeld.
- Banton, O., Laroque, M. 1997. *Agriflux 2.0. Manual d'utilisation. Logiciel d'évaluation des pertes environnementales de nitrates et pesticides agricoles*, INRS-Eau, Université du Quebec, 145pp.
- Bartlett, M.S. 1949. Fitting a straight line when both variables are subject to error. *Biometrics* 5: 207-212.
- Beaton, A.E., Tukey, J.W. 1974. The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics* 16: 147-185.
- Benne, I., Heineke, H.J., Nettelmann, R. 1990. Die DV-gestützte Auswertung der Bodenschätzung. Erfassungsanweisung und Übersetzungsschlüssel. Technische Berichte zum NIBIS. Schweizerbart'sche Verlagsbuchhandlung Stuttgart.
- Beran, R. 1974. Asymptotically efficient adaptive rank estimates in location models. *Ann. Statist.* 2: 67-113.
- Beudert, G. 1997. Gewässerbelastung und Stoffaustrag von befestigten Flächen in einem kleinen ländlichen Einzugsgebiet. Schriftenreihe des Instituts für Siedlungswasserwirtschaft der Universität Karlsruhe Band 80. Kommissionsverlag R. Oldenbourg.
- Bienwald, G. 1970. Menzingen, ein Gang durch 1200 Jahre Geschichte. Hrsg.: Gemeinde Menzingen.
- Birkinshaw, S.J., Ewen, J. 2000a. Nitrogen transformation component for SHETRAN catchment nitrate transport modelling. *Journal of hydrology* 230 (1-2): 1-17.
- Birkinshaw, S.J., Ewen, J. 2000b. Modelling nitrate transport in the Slapton Wood catchment using SHETRAN. *Journal of hydrology* 230 (1-2): 18-33.
- Bodendiek, R. Henn, R. 1990. *Topics in Combinatorics and Graph Theorie*. Verlag Physika Heidelberg.
- Bork, H.-R., Ricken, W. 1983. Bodenerosion, holozäne und pleistozäne Bodenentwicklung. *Catena/Supplement*, Band 3.
- Breiman, L., Friedman, J.H., Olshen, R. A., Stone, C.J., 1984. *Classification and regression trees*. Pacific Grove Wadsworth.
- Breiman, L. 1996a. Bagging predictors. *Machine Learning* 24: (2) 123-140.
- Breiman, L. 1996b. Technical note: Some properties of splitting criteria. *Machine Learning* 24: (1) 41-47.
- Breiman, L. 1998a. Arcing classifiers. *Ann. Stat.* 26: (3) 801-824.
- Breiman L. 1998b. Arcing classifiers – Rejoinder. *Ann. Stat.* 26: (3) 841-849.

- Bronstert, A. 1994. Modellierung der Abflußbildung und der Bodenwasserdynamik von Hängen. Mitteilungen des Instituts für Hydrologie und Wasserwirtschaft, Universität Karlsruhe. Band 46.
- Buck, W. 1994. Stofftransport in kleinen ländlichen Einzugsgebieten: Verbundprojekt Weiherbach in: Niederschlag-Abfluss-Modelle für kleine Einzugsgebiete und ihre Anwendung. 18. Fortbildungslehrgang Hydrologie 19.-23 September 1994 in Karlsruhe des Deutschen Verbandes für Wasserwirtschaft und Kulturbau e.V. DVWK.
- Bünig, H. 1991. Robuste und adaptive Tests. 339 S. de Gruyter Verlag, Berlin.
- Cappelle, A., Ahlers, E. 1998. Zum Abgleich der Ansprache von Bodenprofilen nach dem Bodenschätzungs-gesetz und der Kartieranweisung des Niedersächsischen Landesamtes für Bodenforschung. Mitteilungen der Deutschen Bodenkundlichen Gesellschaft 88: 281-284.
- Chipman, H.A., George, E.I., McCulloch, R.E., 1998. Bayesian CART model search. Journal of American Statistical Association. 93: (443) 935-948.
- Chipman, H.A., George, E.I., McCulloch, R.E. 2000. Hierarchical priors for bayesian CART shrinkage. Stat Comput 10: (1) 17-24.
- Clark, W.R., Schmitz, R.A., Bogenschutz, T.R. 1999. Site selection and nest success of ring-necked pheasants as a function of location in Iowa landscapes. J. Wildlife Manage 63: (3) 976-989.
- Cochran, W.G. 1947. Consequences when the Assumptions for the Analysis of Variance are not satisfied. Biometrics 3 (1): 22-38
- Comfort, S.D., Inskip, W.P., Lockerman, R.H. 1993. Observed and simulated transport of a conservative tracer under line-source irrigation Journal of Environmental Quality. 22 (3): 554-561.
- Cooper, D.M., Ragab, R. Lewis, D.R., Whitehead, P.G. 1994. Modelling nitrate leaching to surface waters. Report for MAFF/NERC, Institute of Hydrology, Wallingford, Oxfordshire, UK.
- Corwin, D.L., Goldberg, S., David, A. 1999. Evaluation of a functional model for simulating boron transport in soil. Soil Science 164 (10): 697-717.
- Cox, D.R., Oakes, D. 1984. Analysis of Survival Data. Chapman and Hall. London, New York.
- Cox, D.R., Hinkley, D.V. 1974. Theoretical Statistics. London, Chapman and Hall.
- Daughtry, C.S.T., Walthall, C.L., Kim, M.S., de Colstoun, E.B., McMurtrey, J.E. 2000. Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. REMOTE SENSING OF ENVIRONMENT. 74 (2): 229-239.
- Davis, R.E., Elder, K., Howlett, D., Bouzaglou, E. 1999. Relating storm and weather factors to dry slab avalanche activity at Alta, Utah, and Mammoth Mountain, California, using classification and regression trees. Cold Reg. Sci. Technol. 30: (1-3) 79-89.
- Daniel, C. 1976. Applications of statistics to Industrial Experimentation. New York. Wiley- Interscience.
- Daniel, C., Wood, F.S. 1971. Fitting equationa to data. Computer analysis of multifactor data for Scientists an Engineers. New York: Wiley- Interscience.
- DeFries, R., Hansen, M., Steininger, M., Dubayah, R., Sohlberg, R., Townshend, J. 1997. Subpixel forest cover in central Africa from multisensor, multitemporal data. Remote Sensing of Environment 60: (3) 228-246.
- De Fries, R.S., Hansen, M., Townshend, J.R.G., Sohlberg, R. 1998. Global land cover classifications at 8 km spatial resolution: the use of training data derived from Landsat imagery in decision tree classifiers. International Journal of Remote Sensing 19: (16) 3141-3168.
- Delgado, J.A., Ristau, R.J., Dillon, M.A., Duke, H.R., Stuebe, A., Follett, R.F., Shaffer, M.J., Riggerbach, R.R., Sparks R.T., Thompson, A., Kawanabe, L.M., Kunugi, A., Thompson, K. 2001. Use of innovative tools to increase nitrogen use efficiency and protect environmental quality in crop rotations. COMMUNICATIONS IN SOIL SCIENCE AND PLANT ANALYSIS. 32 (7-8): 1321-1354.
- Delbrück, M. 1997. Großflächiges Bromid-Tracerexperiment zur zeitlichen und räumlichen Variabilität des Gewässertransports an einem Lösshang. Dissertation, Institut für Umweltphysik, Naturwissenschaftlich-Mathematische Gesamtfakultät, Universität Heidelberg.
- Denison, D.G.T., Mallick, B.K., Smith, A.F.M. 1998. A Bayesian CART algorithm. Biometrika 85: (2) 363-377.

- Deutscher Wetterdienst (Hrsg.) 1953. Klimaatlas von Baden-Württemberg. Bad Kissingen.
- Diemann, R. 1998. Auswertung der Bodenschätzung für das Gebiet der neuen Bundesländer. Mitteilungen der Deutschen Bodenkundlichen Gesellschaft 88: 293-296.
- Diercke, C. 1988. Diercke-Weltatlas. Begr.: Carl Diercke. 1. Aufl. d. Neubearb. 275 S. Westermann, Braunschweig.
- Dikau, R. 1986. Experimentelle Untersuchungen zu Oberflächenabfluss und Bodenabtrag von Meßparzellen und landwirtschaftlichen Nutzflächen. Heidelberger Geographische Arbeiten 81. Geographisches Institut der Universität Heidelberg.
- Dobbertin, M., Biging, G.S. 1998. Using the non-parametric classifier CART to model forest tree mortality. Forest Sci 44: (4) 507-516.
- Donigian, Jr., A.S., Rao, P.S.C. 1990. Selection, application, and validation of environmental fate models. In: Proc. Int. Symp. On Water Quality Modeling of Agricultural Non-Point Sources, Part II ed: D.G. DeCoursey: 557-600 Washington, D.C. USDA.
- Efron, B. 1979. Bootstrap methods: another look at the jackknife. Ann. Statist. 7: 1-26.
- Efron, B. 1983. Estimating the error rate of a prediction rule: improvements on cross-validation. J. Amer. Statist. Assoc. 78: 316-331.
- Efron, B., Tibshirani, R. J. 1993. An Introduction to the Bootstrap. Monographs on Statistics and applied Probability No. 57. Chapman & Hall, Inc. New York, London.
- Efron B, Tibshirani R.J. 1997. Improvements on cross-validation: The 0.632+ bootstrap method. Journal of the American Statistical Association 92: (438) 548-560.
- Eitel, B. 1989. Morphogenese des südlichen Kraichgau unter besonderer Berücksichtigung tertiärer und pleistozäner Decksedimente. Stuttgarter Geographische Studien 111.
- Emmons, E.E., Jennings, M.J., Edwards, C. 1999. An alternative classification method for northern Wisconsin lakes. Can j fish Aquat Sci 56: (4) 661-669.
- Engel, T., Klöcking, B., Priesack, E., Schaaf, T. 1993. Simulationsmodelle zur Stickstoffdynamik – Analyse und Vergleich, Agrarinformatik Bd 25, Eugen Ulmer-Verlag Stuttgart.
- Engel, E. 1998. Aufgaben des Bundesfinanzministeriums bei der Bodenschätzung. Mitteilungen der Deutschen Bodenkundlichen Gesellschaft 88. 249-252.
- Faucher, M., Burrows, W.R., Pandolfo, L. 1999. Empirical-statistical reconstruction of surface marine winds along the western coast of Canada. Climate Res. 11: (3) 173-190.
- Ferschl, F. 1985. Deskriptive Statistik. 3. korrigierte Auflage. Physica Verlag. Würzburg.
- Fetzer, K.D., Schlicker, P. 1992. Die Rolle der Bodenschätzung im saarländischen Bodeninformationssystem (Saar-BIS). Mitteilungen der deutschen bodenkundlichen Gesellschaft 67: 205-208.
- Firbas, F. 1949. Spät- und nacheiszeitliche Waldgeschichte Mitteleuropas nördlich der Alpen. Fischer-Verlag, Jena.
- Fischer, G. 1992. Aufgaben und Auswertung der Bodenschätzung im ehemaligen Bezirk Erfurt von 1955-1989. Mitteilungen der Deutschen Bodenkundlichen Gesellschaft 67: 209-212.
- Fisher, N.I. 1990. On the application of some computer-intensive statistical methods in the earth sciences. In: Hogarth, W.L., Noye B.J. (Hrsg.) Computational techniques and applications: CTAC-89. Proceedings of the computational techniques and applications conference, Griffith, 1989. Hemisphere, New York: 33-48.
- Fleck, W. 1994. Ermittlung von K-Faktoren – Reichsbodenschätzung und Bodenbestandsaufnahme. In: Prognose von Bodenerosion: Workshop. Materialien zum Bodenschutz. 4. Landesanstalt für Umweltschutz Baden-Württemberg, Karlsruhe: 56-67.
- Fleischmann, R., Hacker, E., Oelkers, K.H.. 1979. Vorschlag zu einem Übersetzungsschlüssel für die automatische bodenkundliche Auswertung der Bodenschätzung. Geologisches Jahrbuch, Reihe F, 6: 3-28.
- Fox, J., Long, J.S. 1990. Modern Methods of Data Analysis. Beverly Hills, London.
- Franchini, M., Wendling, J., Obled, CH., Todini, E. 1996. Physical interpretation and sensitivity analysis of the Topmodel. Journal of Hydrology 175: 293-338.
- Freeman, D.H. Jr. 1987. Applied Categorical Data Analysis. Statistics. Vol. 79. M. Dekker New York, Basel.

- Freund, K.L. 1992. Grundlagen, Ziele und Durchführung der Bodenschätzung. In: Mitteilungen der Deutschen Bodenkundlichen Gesellschaft 67: 213-219.
- Freund, Y., Schapire, R.E. 1997. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences* 55: (1) 119-139.
- Frydman, H., Altman, E.I., Kao, D. 1985. Introducing recursive partitioning for financial classification: the case of financial distress. *The Journal of Finance* 40: 269-291.
- Geologisches Landesamt Baden-Württemberg 1993. Geologische Karte von Baden-Württemberg, Maßstab 1:25.000. Blatt 6818 Kraichtal (früher Odenheim). Bearb. von H. Thürach. Geologisches Landesamt Baden-Württemberg, Stuttgart.
- Gerlinger, K. 1996. Untersuchung der Erosionsneigung von Lössböden als Grundlage der Modellierung des Stofftransports.. Mitteilungen des Instituts für Wasserbau und Kulturtechnik, Universität Karlsruhe.
- Gerold, G., Reum, R., Wagner, S. 1992. Flächenhafte Erfassung der Bodenparameter, Bodendifferenzierung und Bodenerosion. In: E.J. Plate (Hrsg.). Prognosemodell für die Gewässerbelastung durch Stofftransport aus einem kleinen ländlichen Einzugsgebiet; Schlussbericht zur 1. Phase. Institut für Hydrologie und Wasserwirtschaft, Universität Karlsruhe.
- Geyer, O.F., Gwinner, N.P. 1991. Geologie von Baden-Württemberg. 4. Auflage. Schweizerbart'sche Verlagsbuchhandlung, Stuttgart.
- Göbel, B., Kaske, A., Reiche, E.-W., Schleuß, U. 1995. Erstellung einer Konzeptbodenkarte mit Hilfe eines GIS und deren Verifizierung durch eine Kartierung. Mitteilungen der Deutschen Bodenkundlichen Gesellschaft 76: 1045-1048.
- Gölz-Huwe, H., Simon, W., Huwe, B., Van der Ploeg, R.R. 1989. Zum jahreszeitlichen Nitratgehalt und zur Nitratauswaschung von landwirtschaftlich genutzten Böden in Baden-Württemberg. *Z. Pflanzenernährung und Bodenkunde* 152 (1989): 273-280.
- Green, P.J. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82: (4) 711-732.
- Griffin, W.L., Fisher, N.I., Friedman, J.H., Ryan, C.G. 1997. Statistical techniques for the classification of chromites in diamond exploration samples. *Journal of Geochemical Exploration* 59: 233-249.
- Gündra, H. 1992. Untersuchungen zu Relief und Bodenverbreitung im Einzugsgebiet des Biddersbach, Nord-Kraichgau. Diplomarbeit am Geographischen Institut der Universität Heidelberg.
- Guttek, U., Wiedenroth, W., Wallbaum, E. 1998. Projekt „Digitalisierung von Altdaten der Bodenschätzung“ im Land Sachsen-Anhalt. Mitteilungen der Deutschen Bodenkundlichen Gesellschaft 88: 307-310.
- Hampel, F.R. 1968. Contributions to the theory of robust estimation. Ph.D. Thesis. University of California. Berkeley, California.
- Hampel, F.R. 1971. A general qualitative definition of robustness. *Ann. Math. Statist.* 42: 1887-1996.
- Hampel, F.R. 1974. The influence curve and its role in robust estimation. *J. Am. Statist. Assoc.* 69: 383-393.
- Hampel, F. 1980. Robuste Schätzungen: ein anwendungsorientierter Überblick. *Biom. J.* Vol. 22 (1): 3-21.
- Hampel, F.R., Ronchetti, E. M., Rousseeuw, P., J., Stahel, W. A. 1986. Robust statistics. John Wiley & Sons, New York.
- Hansen, M., Dubayah, R., DeFries, R. 1996. Classification trees: An alternative to traditional land cover classifiers. *International Journal of Remote Sensing*. 17: (5) 1075-1081.
- Hansen, M.C., Defries, R.S., Townshend, J.R.G., Sohlberg, R. 2000. Global land cover classification at 1km spatial resolution using a classification tree approach. *International Journal of Remote Sensing*. 21: (6-7) 1331-1364.
- Harrach, T. 1982. Ertragsfähigkeit erodierter Böden – Arbeiten der DLG. 174 Bodenerosion 84-95 DLG-Verlag Frankfurt/M.
- Harrach, T. 1987 Bodenbewertung für die Landwirtschaft und den Naturschutz. *Zeitschrift für Kulturtechnik und Flurbereinigung* 28: 184-190.
- Harrach, T. 1998. Nutzung der Bodenschätzungsdaten für nichtsteuerliche Zwecke – Chancen und Grenzen in Mittelgebirgslandschaften - Mitteilungen der Deutschen Bodenkundlichen Gesellschaft 88: 265-268.

- Hartmann, K.-J., Finner, J., Cordsen, E. 1998. Verfahrensvergleich zur Bewertung der Bodenfunktion „Ausgleichskörper im Wasserkreislauf“ auf Basis der Bodenschätzungsdaten. *Mitteilungen der Deutschen Bodenkundlichen Gesellschaft* 88: 311-314.
- Hartung, J., Elpelt, B., Klösener, K.-H. 1986. *Statistik Lehr- und Handbuch der angewandten Statistik*, 5. Auflage. Oldenbourg, München.
- Heikkonen, J., Varfis, A. 1998. Land cover land use classification of urban areas: A remote sensing approach. *Int. J. Pattern Recogn.* 12: (4) 475-489.
- Henley, W.E., Hand, D.J. 1996. A k-nearest-neighbour classifier for assessing consumer credit risk. *Statistician* 45: (1) 77-95.
- Hennings, V. 1991. Die Bedeutung der räumlichen Variabilität bodenkundlicher Basisdaten für aktuelle und zukünftige Kartiertechniken, dargestellt an einem Beispielgebiet im nördlichen Harzvorland. *Geologisches Jahrbuch Reihe F*, 28: 3-143.
- Henningsen, D., Katzung, G. 1997. *Einführung in die Geologie Deutschlands*. Dt. Taschenbuch Verlag, München 4. Auflage.
- Hierold, W. 1992. Erfahrungen bei der Nutzung von Bodenschätzungsdaten für Bodenkartierung und Bodenbewertung auf Standorten verschiedener Entstehungsarten. *Mitteilungen der Deutschen Bodenkundlichen Gesellschaft* 67: 221-224.
- Hötzl, H., Bohleber, A., Sass, I., Schmaland, G., Wohnlich, S. 1992. Transport bei inhomogener Sickerwasserbewegung. In: E.J. Plate (Hrsg.). *Prognosemodell für die Gewässerbelastung durch Stofftransport aus einem kleinen ländlichen Einzugsgebiet; Schlussbericht zur 1. Phase*. Institut für Hydrologie und Wasserwirtschaft, Universität Karlsruhe: 463-484.
- Hötzl, H., Ackermann, M. 1996. Grundwasserumsatzanalyse des Weiherbachgebietes. Beitrag zur Zwischenpräsentation des Weiherbachprojektes am 10. Dezember 1996 in Karlsruhe. Institut für Hydrologie und Wasserwirtschaft, Universität Karlsruhe. Hötzl, H., Ackermann, M. 1998. Schlussbericht des Teilprojekts 13 des Weiherbachprojektes. In: Plate, E.J. (Hrsg.): *Prognosemodell für die Gewässerbelastung durch Stofftransport aus einem kleinen ländlichen Einzugsgebiet*.
- Hogg, R.V. 1974. Adaptive robust procedures: A partial review and some suggestions for future applications and theory (with discussion). *J. Amer. Statist. Ass* 69: 909-927.
- Holland, P.W., Welsch, R.E. 1977. Robust regression using iteratively reweighted least-squares. *Commun. Statist. A* 6: 813-888.
- Hoffmann, G. 1991. *Die Untersuchung von Böden. Methodenbuch Band I*. VDLUFA-Verlag, Darmstadt.
- Hoffmann, K. 1993. Modellierung zu Boden-Reliefbeziehungen im Lösshügelland, Süd-Kraichgau. Diplomarbeit am Geographischen Institut der Universität Heidelberg.
- Hoffmann, K., Fleck, W., Gündra, H., Dikau, R. 1993. Computergestützte Modellierung zu Relief-Beziehungen in Lössgebieten Nord-Baden-Württembergs. *Mitteilungen der Deutschen Bodenkundlichen Gesellschaft* 72: 935-938.
- Hofmann, H.-J. 1990. Die Anwendung des CART-Verfahrens zur statistischen Bonitätsanalyse von Konsumtenkrediten. *Zeitschrift für Betriebswirtschaft* 60: 941-962.
- Howieson, B. 1991. A security analyst's action recommendations: an application of recursive partitioning to modelling judgement. *Australian Journal of Management* 16: 165-185.
- Huber, P.J. 1964. Robust estimation of a location parameter. *Ann. Math. Statist.* 35: 73-101.
- Huber, P.J. 1981. *Robust Statistics*. Wiley. New York.
- Huwe, B. 1992b. WHNSIM Programm-Dokumentation. Version 2.0.
- Huwe, B., Totsche, K.U. 1995. Deterministic and stochastic modelling of water, heat and nitrogen dynamics on different scales with WHNSIM. *Journal of contaminant hydrology*. 20 (3-4): 265-284.
- Huwe, B., Schillinger, M.P., Schulz, K. 1997. Modellierung der flächigen Stickstoffdynamik eines überwiegend landwirtschaftlich genutzten Einzugsgebietes. Schlussbericht Teilprojekt 17. In BMBF-Verbundvorhaben: *Prognosemodell für die Gewässerbelastung durch Stofftransport aus einem kleinen landwirtschaftlich genutzten Einzugsgebiet*. (Hrsg. E.J.Plate) Lehrstuhl Bodenkunde Abteilung Bodenphysik, Universität Bayreuth.

- Ilseman, J., Goeb, S., Bachmann, J. 2001. How many soil samples are necessary to obtain a reliable estimate of mean nitrate concentrations in an agricultural field? *J. Plant Nutr. Soil Sc.* 164 (5): 585-590.
- Jang, J.S.R., Sun, C.T. 1995. Neuro-fuzzy modeling and control. *Proceedings of the IEEE*. 83: (3) 378-406.
- Janssen, P.H.M., Heuberger, P.S.C. 1995. Calibration of process orientated models. *Eco. Model.* 83: 55-66.
- Johns, M.V. 1979. Robust Pitman-like estimators. in: *Robustness of Statistics*, Hrsg. R.L. Launer und G.N. Wilkinson. Academic Press, N.Y.: 49-60.
- Kallianpur, G., Rao, C.R. 1955. On Fisher's lower bound to asymptotic variance of a consistent estimate. *Sankhyā*. 15: 331-342.
- Kern, U. 2000. mdl. Mitteilung, Sachbearbeiter Landwirtschaft bei der Oberfinanzdirektion Chemnitz; Tel. (03 41) 3 47 – 34 01.
- Kersebaum, K.C., Richter, J., Utermann, J. 1987. Die Simulation der Stickstoffdynamik von Ackerböden unter Getreidevegetation. *Mitteilg. Dtsch. Bodenkundl. Gesellsch.* 55/II:613-618.
- Kiselev, M.V., Ananyan, S.M., Arseniev, S.B. 1997. Regression-based classification methods and their comparison with decision tree algorithms. *Lect Notes Artif Int.* 1263: 134-144.
- Kleefisch B., Köthe, R. 1993. Wege zur rechnergestützten bodenkundlichen Interpretation digitaler Reliefdaten. *Geologisches Jahrbuch, Reihe F*, 27: 59-122.
- Kleefisch, B. 1993. Die automatisierte Vorhersage der Bodenverbreitung – eine Anwendung für wissensbasierte Systeme?. *Geologisches Jahrbuch, Reihe F*, 27: 59-122.
- Kleinbaum, D.G., Kupper, L.L., Chambless, L.E. 1982. Logistic Regression Analysis of Epidemiologic Data. Theory and Praxis. *Communications in Statistics – Theory and Methods* 11: 485 - 547.
- Kohonen, T. 1984. Self-organization and associative memory. Springer Verlag Berlin.
- Kolcz, A. 2000. N-tuple network, CART, and bagging. *Neural. Comput.* 12: (2) 293-304.
- Kolcz, A., Allinson, N.M. 1996. N-tuple regression network. *Neural Networks* 9: (5) 855-869.
- Kolle, O., Fiedler, F. (1999). Schlussbericht des Teilprojektes 2 des Weiherbachprojektes, in: Plate E.J. (Hrsg.): Prognosemodell für die Gewässerbelastung durch Stofftransport in einem kleinen ländlichen Einzugsgebiet. Schlussbericht.
- Kraft, M. 1998. GIS-gestützte Modellierung einer Bodenformenkarte mit dem Klassifikationsverfahren CART unter besonderer Berücksichtigung der Bodenschätzung. Diplomarbeit am Lehrstuhl für Bodenkunde und Bodengeographie, Abteilung Bodenphysik der Universität Bayreuth.
- Kühner, C. 1998. Stand der Digitalisierung der Bodenschätzungsergebnisse in Deutschland. *Mitteilungen der Deutschen Bodenkundlichen Gesellschaft* 88: 253-256.
- Kuhnert, P.M., Do, K.A., McClure, R. 2000. Combining non-parametric models with logistic regression: an application to motor vehicle injury data. *Comput. Stat. Data An.* 34: (3) 371-386.
- Lachenbruch, P.A. 1975. Discriminant Analysis. Hafner Press. New York.
- Lamon, E.C., Carpenter, S.R., Stow, C.A. 1999a. Rates of decrease of polychlorinated biphenyl concentrations in five species of Lake Michigan salmonids. *Can. J. Fish. Aquat. Sci.* 56(1): 53-59.
- Lamon, E.C., Stow, C.A. 1999b. Sources of variability in microcontaminant data for Lake Michigan salmonids: Statistical models and implications for trend detection. *Can. J. Fish. Aquat. Sci.* 56(1): 71-85.
- Lamp, J. 1998. Auswertung des Erosionsgrades und –Trends von Bodenlandschaften aus Grablochdaten. *Mitteilungen der deutschen Bodenkundlichen Gesellschaft* 88, 289-292.
- Landesvermessungsamt Baden-Württemberg 1996. Topographische Karte 1:25.000. Blatt 6818 Kraichtal. Freiburg i. Br.
- Larsen, P. 1994. Untersuchung über die räumliche und zeitliche Variabilität der Erosionsneigung von Böden als Grundlage eines Modells zur Erfassung der Gewässerbelastung durch Stofftransport durch Feststoffeintrag. In: Zwischenbericht (Stand 31.12.1993) des BMFT-Verbundprojekts Prognosemodell für die Gewässerbelastung durch Stofftransport in einem kleinen ländlichen Einzugsgebiet (Weiherbachprojekt). Karlsruhe: 71-78.
- Lasserre, F., Razack, M., Banton, O. 1999. A GIS-linked model for the assessment of nitrate contamination in groundwater. *Journal of hydrology*. 224 (3-4): 81-90.

- Lax, D.A. 1975. An interim report of a Monte Carlo study of robust estimators of width. Technical Report 93, Series 2. Departement of Statistics, Princeton University. Princeton, N.Y.
- LeBlanc, M., Tibshirani, R. 1998. Monotone shrinkage of trees. *J. of Computational and Graphical Stat.* 7 (4): 417-433.
- Lehmann, W. 1995. Anwendung geostatistischer Verfahren auf die Bodenfeuchte in ländlichen Einzugsgebieten. Institut für Hydrologie und äwasserwirtschaft Universität Karlsruhe, 52.
- Lieberoth, J. 1969. Bodenkundliche Bodenfruchtbarkeit. 2. Auflage. Deutscher Landwirtschaftsverlag. Berlin.
- Lienert Gustav A. 1973. Verteilungsfreie Methoden in der Biostatistik/1. 2., völlig neu bearb. Aufl.. Verlag Anton Hain.
- Loague, K.M., Green, R.E. 1991. Statistical and graphical methods for evaluating solute transport models: Overview and application. *J. Contam. Hydrol.* 7 (1/2): 51-32.
- Lotka, H.J. 1925. Elements of Physical Biology. Williams and Wilkins, Baltimore
- Lunn, R.J., Adams, R., Mackay, R., Dunn, S.M. 1996. Development and application of a nitrogen modelling system for large catchments. *Journal of hydrology.* 174 (3-4): 285-304.
- Mair J., Smidt, J., Lechleitner, P., Dienstel, F., Puschendorf, B. 1995. A decision tree for the early diagnosis of acute Myocardial-Infarction in nontraumatic chest patients at hospital admission. *Chest.* 108(6): 1502-1509.
- Manoukian, E.B. 1985. Modern concepts and Theorems of mathematical statistics. Springer-Verlag New York.
- Marais, M.L., Patell, J.M., Wolfson, M.A. 1984. The experimental design of classification models: an application of recursive partitioning and bootstrapping to commercial bank loan classifications. *Supplement to the Journal of Accounting Research* 22: 87-118.
- Mauder, S. 1992. Geologische, sedimentologische und hydrogeologische Untersuchungen der flachen Grundwasserverhältnisse des Weiherbachtals/Kraichgau. Diplomarbeit am Lehrstuhl für angewandte Geologie der Universität Karlsruhe.
- McBratney, A., Odeh, I. 1997. Application of fuzzy sets in soil science: fuzzy logic, fuzzy measurements and fuzzy decisions. *Geoderma* 77: 85-113.
- McBratney, A.B., Odeh, I.O.A., Bishop, T.F.A., et al. 2000. An overview of pedometric techniques for use in soil survey. *Geoderma* 97(3-4): 293-327.
- McKenzie, N.J., Ryan, P.J. 1999. Spatial prediction of soil properties using environmental correlation. *Geoderma* 89(1-2): 67-94.
- Merler, S., Furlanello, C. 1997. Selection of tree-based classifiers with the bootstrap 0.632+ rule. *Biometrical J.* 39(3): 369-382.
- Mertens, M. 2000. Die Anwendung der Fuzzy-Set-Theorie auf die Modellierung von Stickstoffbilanzen im Weißenstädter Becken. Doktorarbeit an der Universität Bayreuth, Lehrstuhl Bodenkunde Abteilung Bodenphysik.
- Michel, O.J.J., Hero, A.O., Badel, A.E. 1999. Tree-structured nonlinear signal modeling and prediction. *IEEE Transactions On Signal Processing* 47(11): 3027-3041.
- Miller, R. 1964. A trustworthy jackknife. *Ann. Math. Statist.* 35: 1594-1605.
- Miller, R. 1974. The jackknife - A review. *Biometrika.* 61: 1-15
- Morgan, J.N., Sonquist, J.A. 1963. Problems in the analysis of survey data, and a proposal. *J. Amer. Statist. Assoc.* 58: 415-434.
- Morishita, S. Nakaya, A., 1999. Discovery science, proceedings lecture notes in artificial intelligence. 1721. 345-346.
- Morishita, S., Nakaya, A. 2000. Expressive tests for classification and regression. *IEICE T. Inf. Syst.* E83D (1): 52-60.
- Mosteller, F., Tukey, J.W. 1977. Data Analysis and Regression. Addison-Wesley. Reading Mass.
- Nagel, M., Wernecke, K.-D., Fleischer, W. 1994. Computergestützte Datenanalyse. Verlag Technik GmbH Berlin.

- Nash, J.E., Sutcliffe, J.V. 1970. River flow forecasting through conceptual models. Part I. A discussion of principles. *Journal of Hydrology* 10 (3).
- Negron, J.F. 1998. Probability of infestation and extent of mortality associated with the Douglas-fir beetle in the Colorado Front Range. *Forest Ecol. Col. Manag* 107(1-3): 71-85.
- Nestler, I. 1999. GIS-gestützte Regionalisierung von Bodenprofilen mit dem Klassifikationsverfahren CART. unveröffentlichte Diplomarbeit am Lehrstuhl für Bodenkunde und Bodengeographie, Abteilung Bodenphysik, Universität Bayreuth.
- Oelkers, K.H. 1993a. Führung der Bodenschätzungsdaten beim Niedersächsischen Landesamt für Bodenforschung. In: Bodenschätzung und Niedersachsen heute. *Nachrichten der Niedersächsischen Vermessungs- und Katasterverwaltung* 4: 242-261.
- Oelkers, K.H. 1993b. Thematische Auswertung der Bodenschätzung mit dem Niedersächsischen Bodeninformationssystem NIBIS. In: Bodenschätzung in Niedersachsen heute. *Nachrichten der Niedersächsischen Vermessungs- und Katasterverwaltung* 4: 262-269.
- Ostmann, U. 1993. Die Übernahme der historischen Landnutzungskarten aus Karten des 18. und 19. Jahrhunderts in das Niedersächsische Bodeninformationssystem. *Geologisches Jahrbuch, Reihe F*, 27: 145-183.
- Parish, R.S., Smith, C.N. 1990. A method for testing whether model predictions fall within a prescribed factor of true values, with an application to pesticide leaching. *Ecological Modelling* 51 (1/2): 59-72.
- Pécsi, M., Richter, G. 1996. Löss: Herkunft – Gliederung – Landschaften. *Zeitschrift für Geomorphologie, Supplement* 98. Bornträger-Verlag Berlin, Stuttgart.
- Pennel, K.D., Hornsby, A.G., Jessup, R.E., Rao, P.S.C. 1990. Evaluation of five simulation models for predicting aldicarb and bromide behavior under field conditions. *Water Resources Research* 26 (11): 2679-2693.
- Peter, M., Siebert, S. 1998. Nutzung der Bodenschätzung in der hessischen Praxis des Boden- und Gewässerschutzes – Möglichkeiten und Grenzen. *Mitteilungen der Deutschen Bodenkundlichen Gesellschaft* 88: 273-276.
- Pfeiffer, E.-M., Keil, B., Altermann, M. 1998. Nutzung der Bodenschätzung zur Bewertung von Bodenfunktionen und für die Bodenschutzplanungen. – Eine erste Bilanz zur Fachtagung im September 1998 in Frankfurt/Main. *Mitteilungen der Deutschen Bodenkundlichen Gesellschaft* 88: 353-356.
- Pflug, R. 1982. Bau und Entwicklung des Oberrheingrabens. *Erträge der Forschung* Nr. 184. Wissenschaftliche Buchgesellschaft, Darmstadt.
- Pitman, E.J.G. 1949. *Lecture Notes on Nonparametric Statistics*. Columbia University, New York.
- Plant, R.E., Mermer, A., Pettygrove, G.S., Vayssières, M.P., Young, J.A., Miller, R.O., Jackson, L.F., Denison, R.F., Phelps, K. 1999. Factors underlying grain yield spatial variability in three irrigated wheat fields. *Transactiona of the Asae* 42(5): 1187-1202.
- Plate, E.J. 1992. Weiherbachprojekt. „Prognosemodell für die Gewässerbelastung durch Stofftransport aus einem kleinen ländlichen Einzugsgebiet“ Schlussbericht zur 1. Phase des BMBF-Verbundprojektes. Hrsg: E.J. Plate, Institut für Wasserwirtschaft und Hydrologie Universität Karlsruhe (IHW).
- Prohorov, Y.V. 1956. Convergence of random processes and limit theorems in probability theory. *Theor. Prob. Appl.* 1: 157-214.
- Rau, D. 1992. Das Für und Wider zum Konzept der „Reichsbodenschätzung“ vor deren Beginn am Anfang der 30er Jahre. *Mitteilungen der Deutschen bodenkundlichen Gesellschaft* 67: 245-248.
- Refsgaard, J.C., Thorsen, M., Jensen, J.B., Kleeschulte, S., Hansen, S. 1999. Large scale modelling of groundwater contamination from nitrate leaching. *Journal of hydrology* 221 (3-4): 117-140.
- Reichard, S.H., Hamilton, C.W. 1997. Predicting invasions of woody plants introduced into North America. *Conserv. Biol.* 11(1): 193-203.
- Reiche, E. W. 1991. Entwicklung, Validierung und Anwendung eines Modellsystems zur Beschreibung und flächenhaften Bilanzierung der Wasser- und Stickstoffdynamik in Böden. *Kieler Geographische Schriften* Band 79, Selbstverlag des Geographischen Institutes der Universität Kiel.

- Reiche, E.W., Schleuß, U. 1992. Untersuchungen zur Aussagegenauigkeit von Daten der Bodenschätzung anhand der Ergebnisse einer aktuell durchgeführten Bodenkartierung mit Hilfe eines Geographischen Informationssystems (GIS). *Mitteilungen der Deutschen Bodenkundlichen Gesellschaft* 67: 249-252.
- Reiche, E.-W. 1994. Modelling water and nitrogen dynamics on a catchment scale. *Ecol. Model.* 76: 371-384.
- Richter, G.M., Baumann, K., Richter, J. 1996. Simulation des Nitrataustrages im Winterhalbjahr aus intensiv genutzten Ackerböden der Geest. *Z. Pflanzenernährung und Bodenkunde* 159 (1996): 279-288.
- Rinne, H. 1995. *Statistische Methoden der Qualitätssicherung*. 3. überarb. Auflage, München, Wien.
- Rösch, A., Kurandt, F. 1991. *Bodenschätzung, Gesetze mit amtlicher Begründung, Durchführungsbestimmungen und Verwaltungsvorschriften, erläutert von Rösch/Kurandt, z.T. fotomech. Nachdruck der Ausgabe 1950*. 3. Auflage. Heymann Köln, Berlin, Bonn, München, 148 S.
- Rothaar, H., Schröder, D. 1987. Erweiterungsmöglichkeit der Bodenschätzung zur besseren Verwendbarkeit für planerische und bodenkundliche Zwecke. *Mitteilungen der Deutschen Bodenkundlichen Gesellschaft* 53: 249-254.
- Rousseeuw, P.J. 1981. A new infinitesimal approach to robust estimation. *Z. Wahrsch. verw. Geb.* 56: 127-132.
- Rousseeuw, P.J., Yohai, V. 1984. Robust regression by means of S-estimators. in: *Robust and Nonlinear Time Series Analysis*. Hrsg. J. Franke, W. Härdle, R.D. Martin. *Lecture Notes in Statistics* 26. Springer, N.Y.: 256-272.
- Sachs, L. 1997. *Angewandte Statistik*, 8. Auflage. Springer Verlag Berlin/Heidelberg.
- Sacks, J. 1975. An asymptotically efficient sequence of estimators of a location parameter. *Ann. Statist.* 3: 285-298.
- Sauer, S. 1998. Erstellung eines regionalen Übersetzungsschlüssels zur Ableitung von Bodenkennwerten aus Bodenschätzungsdaten. *Mitteilungen der Deutschen Bodenkundlichen Gesellschaft* 88: 285-288.
- Sauer, S. 1999. Nutzung der Bodenschätzung zur Erstellung von Bodenfunktionskarten, dargestellt am Beispiel der nutzbaren Feldkapazität. *Mitteilungen der Deutschen Bodenkundlichen Gesellschaft* 91: 1076-1079.
- Schaar, J. 1989. *Untersuchungen zum Wasserhaushalt kleiner Einzugsgebiete im Elsenzthal/Kraichgau*. Heidelberger Geographische Arbeiten Nr. 86. Geographisches Institut der Universität Heidelberg.
- Schäfer, D. 1999. *Bodenhydraulische Eigenschaften eines Kleineinzugsgebietes – Vergleich und Bewertung unterschiedlicher Verfahren*. Dissertationsreihe am Institut für Hydromechanik der Universität Karlsruhe. Heft 1999/1.
- Schmidhalter, U., Strasser, R., Oertli, J.J. 1992. Zeitliche und räumliche Variabilität von Mineralstickstoff im Boden. *Landwirtschaft Schweiz* 5 (7), 354-362.
- Schmidhalter, U., Alföldi, T., Oertli, J.J. 1991. Repräsentativität von N_{min} -Untersuchungen. *Landwirtschaft Schweiz* 4. 431-435.
- Schottmüller, H. 1961. Der Löß als gestaltender Faktor in der Kulturlandschaft des Kraichgaus, *Forschungen zur deutschen Landeskunde*. Bd 130. Arbeit aus dem geographischen Institut der Technischen Hochschule Karlsruhe. Bad Godesberg.
- Schulla, J. 1997. Hydrologische Modellierung von Flussgebieten zur Abschätzung der Folgen von Klimaänderungen. *Zürcher Geographische Schriften* Heft 69. Geographisches Institut ETH.
- Schulte, A. 1995. Hochwasserabfluss, Sedimenttransport und Gerinnebettgestaltung an der Elsenz im Kraichgau. *Heidelberger Geographische Arbeiten* 98. Geographisches Institut der Universität Heidelberg.
- Schweizer, V., Kraatz, R. 1982. *Kraichgau und südlicher Odenwald.. Sammlung geologischer Führer* 72, Bornträger, Berlin.
- Siciliano, R., Mola, F. 2000. Multivariate data analysis and modeling through classification and regression trees. *Comput Stat Data An.* 32(3-4): 285-301.
- Simunek, M., Sejna, M., van Genuchten, M. 1999. *HYDRUS 2D. User Manual, Version 2.0*. International Ground Water Modelling Center, Colorado School of Mines: 227.
- Sonquist, J.A., Baker, E.L., Morgan, J.N. 1964. The detection of interaction effects. Institute for Social Research, University of Michigan. Ann Arbor.

- Srinivasan, V., Kim, Y.H. 1987. Credit Granting: A comparative Analysis of Classification Procedures. *The Journal of Finance* 42: 665-683.
- Stark, H., Göbel, B., Reiche, E., Widmoser, P. 1997. Darstellung und Bewertung von mesoskaligen Stickstoffmodellen. Bundesanstalt für Gewässerkunde, Berlin.
- Stein, C. 1957. Efficient nonparametric testing and estimation. in: *Proc. Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1. University of California Press. Berkeley.
- Steinberg, D., Colla, P. 1997. CART: Classification and Regression Trees. Treestructured Non-Parametric Data Analysis. Interface and Documentation. Salford Systems, San Diego.
- Stevens, S.S. 1946. On the theory of scales of measurement. *Science* 103: 677-680.
- Stigler, S.M. 1980. Studies in the history of probability and statistics XXXVIII: R.H. Smith, a victorian interested in robustness. *Biometrika* 67: 217-221.
- Stone, C.J. 1975. Adaptive maximum likelihood estimators of a location parameter. *Ann Statist.* 3: 267-284.
- Styczen, M., Storm, B. 1993. Modeling of n-movements on catchment scale - a tool for analysis and decision-making. 1. Model description. *Fertilizer Research* 36 (1): 1-6.
- Suarez, A., Lutsko, J.F. 1999. Globally optimal fuzzy decision trees for classification and regression. *IEEE T. Pattern Anal.* 21(12): 1297-1311.
- Sun, N.Z. 1994. *Inverse Problems in Groundwater Modelling*. Kluwer Academic Publishers, Dordrecht, 337p.
- Tafel, E., Reibnegger, G. 1999. Artificial neural networks in laboratory medicine and medical outcome prediction. *Clin. Chem. Lab. Med.* 37(9): 845-853.
- Takeuchi, K. 1971. A uniformly asymptotically efficient estimator of a location parameter. *J. Amer. Statist. Ass.* 66: 292-301.
- Taylor, P.C., Jones, M.C. 1996. Splitting criteria for regression trees. *J. Stat. Comp. and Simulation* 55(4): 267-285.
- Teufel, M. 1998. Klassifikationsverfahren und Aspekte der Variablenkategorisierung beim Kreditscoring. Diplomarbeit am Lehrstuhl Mathematik VII, Universität Bayreuth.
- Thres, P., Totsche, K. U., Huwe, B. 1998. Simulation der Stickstoffdynamik im Einzugsgebiet des Trinkwassersaltsperrsystems Weila/Zeulenroda. Abschlußbericht für die Thüringer Talsperrenverwaltung.
- Tibshirani, R. 1996. Regression shrinkage and selection via the Lasso. *J. of the Royal Statistical Society. Series B-Methodological* 58(1): 267-288.
- Timmermann, F., Mokry, M., Ritz, D. 1998. Schlussbericht des Teilprojekts 10b des Weiherbach-Projektes. In: Plate E.J. (Hrsg.): *Prognosemodell für die Gewässerbelastung durch Stofftransport in einem kleinen ländlichen Einzugsgebiet; Schlussbericht*.
- Tronstad, R. 1995. Importance of melon type, size, grade, container, and season in determining melon prices. *Journal of agricultural and resource economics* 20(1): 32-48.
- Tukey, J.W. (1970) 1977. *Exploratory Data Analysis*. Addison-Wesley Reading, Mass. (1970: vervielfältigte Vorausgabe).
- Tychon, B., Van der Borgh, P., De Backer, L.W. 1999. Water and nitrogen transfer study through soils of a small agricultural water catchment *WATER SCIENCE AND TECHNOLOGY*. 39 (12): 69-76.
- Vitvar, T. 1998. Water Residence Times and Runoff Generation in a Small Prealpine Catchment. *Zürcher Geographische Schriften Heft 71*. Geographisches Institut ETH.
- Vogel, T., Huang, K., Zhang, R., van Genuchten, M. 1996. The HYDRUS-code for simulating onedimensional water flow, solute transport and heat movement in variably-saturated media, Version 5.0. Research-Report Nr. 140. U.S. Salinity laboratory, USDA, ARS, Riverside.
- Volterra, V. 1926. Variazioni e fluttuazioni del numero d'individui in specie animali conviventi. *Mem. Accad. Lincei* 6: 31-113.
- Wagner, J. 1998. Die Bodenschätzung in Österreich. *Mitteilungen der Deutschen Bodenkundlichen Gesellschaft* 88: 257-260.
- Wagner, M. 1997. Das CART-Verfahren im Rahmen des Kreditscoring. Diplomarbeit am Lehrstuhl Mathematik VII, Universität Bayreuth.

- Wagner, S. 1991. Bodendifferenzierung in einer landwirtschaftlich genutzten Region des Kraichgaus. *Geoökodynamik* 12: 207-235.
- Wallbaum, E. 1991. Die Übersetzung Bodenarten beschreibender Signa aus Gefügebeschrieben der Bodenschätzung in Boden- und Körnungsarten. *Mitteilungen der Deutschen Bodenkundlichen Gesellschaft*. 67: 287-290.
- Wallbaum, E. 1991. Ableitung von Informationen zur Bodenkartierung aus Ergebnissen der Reichsbodenschätzung. Dissertation an der Humboldt-Universität Berlin.
- Wallbaum, E., Blaser, P., Heineke, H.-J., Schultz-Sternberg, R. 1998. Nutzung der Bodenschätzung beim Aufbau von Bodeninformationssystemen in den Ländern. *Mitteilungen der Deutschen Bodenkundlichen Gesellschaft*. 88: 301-306.
- Walmsley, J.L., Burrows, W.R., Schemenauer, R.S. 1999. The use of routine weather observations to calculate liquid water content in summertime high-elevation fog. *J. Appl. Meteorol.* 38(4): 369-384.
- Walter, H. 1986. *Allgemeine Geobotanik als Grundlage einer ganzheitlichen Ökologie* 3. Auflage. Ulmer, Stuttgart.
- Wegehenkel, M. 1996. Anwendung des Agrarökosystemmodells EPIC zur Prognose des Wasser- und Nährstoffhaushaltes von landwirtschaftlich genutzten Flächen auf Sandböden in Ost-Brandenburg. *Z. Pflanzenernährung und Bodenkunde*. 159: 479-490.
- Weiblen, B. 1997. Die Erstellung eines Übersetzungsschlüssels zur Ableitung einer Bodenartenkarte aus den Daten der Bodenschätzung im Einzugsgebiet des Weiherbaches/Kraichgau mit Hilfe des Geographischen Informationssystems ARC/INFO. Diplomarbeit am Lehrstuhl für Bodenkunde und Bodengeographie, Abteilung Bodenphysik, Universität Bayreuth.
- Weltner, K. 1998. Auswertung der Bodenschätzung am Beispiel eines Lössgebietes am Berger Rücken (Main-tal). *Mitteilungen der Deutschen Bodenkundlichen Gesellschaft*. 88: 277-280.
- Wernecke, K.D., Possinger, K., Kalb, G., Stein, J. 1998. Validating classification trees. *Biometrical Journal* 40 (8): 993-1005.
- Whitehead, P.G., Wilson, E.J., Butterfield, D. 1998. A semi-distributed Integrated Nitrogen model for multiple source assessment in Catchments (INCA): Part I - model structure and process equations. *Science of the total environment*. 210 (1-6): 547-558.
- Wilson, A.M. 1998. Software review of CART (TM) for Windows, version 3.1. Human and ecological risk assessment 4(5): 1071-1083.
- Wittmann, O. 1992. Entwicklung und Bedeutung der Bodenschätzung für nichtsteuerliche Zwecke. *Mitteilungen der Deutschen Bodenkundlichen Gesellschaft* 67: 291-296.
- Wolfowitz, J. 1957. The minimum distance method. *Ann. Math. Statist.* 28: 75-88.
- Zacharias, S., Heatwole, C.D., Coakley, C.W. 1996. Robust quantitative techniques for validating pesticide transport models. *TRANSACTIONS OF THE ASAE* 39(1): 47-54.
- Zadeh, L. 1965. Fuzzy sets. *Inf. Control* 8: 338-353.
- Zeller, W. 1992. Zur Schätzung pseudovergleyter Löss-Parabraunerden. *Mitteilungen der Deutschen Bodenkundlichen Gesellschaft*. 67: 297-300.

Anhang 1:

Darstellung der Arbeitsbereiche der einzelnen Institutionen des „Weiherbachprojektes“

Nach Buck (1994), verändert

Teil projekt	Institution	Ergebnisse, Teilmodelle
1 a-c	Institut für Hydrologie und Wasserwirtschaft (IHW), Universität Karlsruhe	Geostatistische Analysen der Bodenfeuchte, Infiltrationsmodell, Federführung des Gesamtprojektes
2	Institut für Meteorologie und Klimaforschung (IMK), Universität Karlsruhe	Meteorologische Größen; Teilmodell Verdunstung
4	Institut für Photogrammetrie und Fernerkundung (IPF), Universität Karlsruhe	Fernerkundungsauswertungen, Raster-GIS
5	Lehrstuhl für Geographie und Geoökologie II, Universität Karlsruhe	Vektor-GIS, flächenhafte und andere Ergebnisdarstellungen
6	Institut für Wasserbau und Kulturtechnik	Teilmodell Feststoffabtrag und -transport
8	Institut für Siedlungswasserwirtschaft (ISWW)	Prozessanalysen für Phosphattransport
10 a,b	Landwirtschaftliche Untersuchungs- und Forschungsanstalt (LUFA) Augustenberg	Bilanzierung/Modellierung des Herbizidtransportes Flächenhafte Darstellung von Vegetationsparametern
13	Institut für Angewandte Geologie (AGK), Universität Karlsruhe	Geologische Kartierung, Bodenparameter, Teilmodelle Makroporenflüsse, Grundwasser
14	Institut für Umweltphysik (IUP), Universität Heidelberg	Grundwasserneubildung, Regionalisierung von Bodenparametern
15	Institut für Hydromechanik (IfH), Universität Karlsruhe	Teilmodell Bodenwasserdynamik
16	Lehrstuhl für Hydrologie und Wasserwirtschaft (LHW), TU Cottbus	Diverse Auswertungen und Teilmodelle
17	Abteilung Bodenphysik, Universität Bayreuth	Bodenkartierung, regionalisierte Boden- und agrarmeteorologische Parameter; Teilmodell Stickstoffdynamik

Anhang 2:

Abgrenzung der Kreuzvalidierung zu Jackknife und Bootstrap

1.1 Einleitung

Das Verfahren der Kreuzvalidierung wird fälschlicherweise oft gleichgesetzt mit dem Jackknife oder dem Bootstrap. Deshalb gehe ich in diesem Kapitel kurz auf die genannten Verfahren ein und zeige Übereinstimmungen, Unterschiede und Verknüpfungspunkte auf.

Die Jackknife- und Bootstrap-Verfahren werden allgemein zur Bestimmung der Genauigkeit von statistischen Maßzahlen (Synonyme nach Sachs, 1997: summarische Größe, Stichprobenfunktion, Statistik) angewendet. Mit den Stichprobenfunktionen werden die Parameter θ von Verteilungen (Sachs, 1997) geschätzt. Die Genauigkeit der Stichprobenfunktionen wird durch Bias, Standardfehler und andere Abweichungsmaße charakterisiert (Efron, 1993). Jackknife und Bootstrap sind Verfahren, die auch dann angewendet werden können, wenn keine Kenntnis über den Verteilungstyp und damit die Verteilungsfunktion vorliegt, d. h. es sind nichtparametrische Methoden zur Parameterschätzung (Sachs, 1997). Die Realisation der mit T_N bezeichneten Schätzfunktion (Synonym: Schätzer) für den Parameter θ wird im Folgenden mit t gekennzeichnet.

Der Vorteil verteilungsunabhängiger Methoden besteht darin, dass die Beschränkung auf die traditionelle parametrische Theorie entfällt und damit die Loslösung von einer kleinen Menge von Standardmodellen möglich ist (Efron, 1983). Die Voraussetzungen der Standardmodelle liegen, streng genommen, in der Praxis nie vor, so dass die parametrischen Methoden meist an den praktischen Erfordernissen vorbeigehen (Sachs, 1997). Außerdem sind verteilungsfreie Methoden auch auf Rangdaten anwendbar (Sachs, 1997). Deshalb sind sie im Vergleich zu parametrischen Verfahren universeller einsetzbar. Sie stellen insofern eine robustere Version der verteilungsabhängigen Parameterschätzverfahren dar, als sie unabhängig von den Verteilungsannahmen sind und deshalb auch gegenüber Verletzungen dieser Annahmen resistent sind. Es gibt jedoch nur eine geringe Überschneidung zwischen robusten und verteilungsfreien Methoden (Huber, 1981). Die Robustheit der parameterfreien Verfahren gegenüber Verteilungsannahmen stellt dabei einen gewichtigen Aspekt dar, der aber nicht dazu verführen soll, verteilungsfreie Methoden mit robusten Methoden gleichzusetzen.

1.2 Bootstrap

Der Bootstrap wurde von Efron 1979 als computergestütztes Verfahren eingeführt, um den Standardfehler einer Schätzung $\hat{\theta}$ eines Parameters θ zu bestimmen. Die Grundidee des Bootstraps ist einfach und schon mindestens zwei Jahrhunderte alt (Efron, 1979).

Bootstrap-Schätzungen werden zur Bestimmung der Genauigkeit von Statistiken (im Folgenden auch Stichprobenfunktionen genannt), die als Schätzungen von Parametern der Grundgesamtheit dienen, herangezogen. Mit dem Bootstrap können Standardfehler und Bias von Stichprobenfunktionen und Konfidenzintervalle für die gesuchten Parameter (Efron, 1993) geschätzt werden. Ebenso ist es möglich, Regressionskoeffizienten von Regressionsmodellen (Efron, 1993), Biaskorrekturen der Stichprobenfunktion und den Vorhersagefehler von Modellen abzuschätzen. Dadurch ist ein Verknüpfungs-

punkt mit dem Verfahren der Kreuzvalidierung gegeben, dessen Zweck die Schätzung des Vorhersagefehlers ist.

Im Folgenden erläutere ich das Bootstrap-Prinzip an Hand der einfachen Bootstrap-Schätzung des Standardfehlers einer Statistik und einer einfachen Bootstrap-Schätzung des Vorhersagefehlers von Modellen (Efron, 1993). Prinzipiell ist der Bootstrap auch bei komplizierteren Zielfunktionen der beiden dargestellten Beispiele anwendbar, deshalb spreche ich beispielhaft den 0.632-Bootstrap-Schätzer zur Schätzung des Vorhersagefehlers von Modellen an, eine entsprechende detailliertere Darstellung liegt jedoch außerhalb des Rahmens dieser Dissertation.

Bei der Bootstrap-Schätzung des Standardfehlers werden zunächst *Bootstrap-Stichproben* (Efron, 1993)

$$\mathbf{x}^{*b} = (x_1^*, x_2^*, \dots, x_n^*) \text{ mit } b = 1, \dots, B \text{ und } x_i^* \in \mathbf{x}, \mathbf{x} = (x_1, x_2, \dots, x_n) \quad (\text{A2.1})$$

vom Umfang n durch n -faches, zufälliges Ziehen mit Zurücklegen vom Originaldatensatz \mathbf{x} gewonnen. Der Originaldatensatz ist ein Vektor, bestehend aus n einfachen Beobachtungen x_i . Beim Ziehen mit Zurücklegen können einige Beobachtungen x_i mehrfach, andere aber nicht in der Bootstrap-Stichprobe enthalten sein. Die Bildung der unabhängigen Bootstrap-Stichproben \mathbf{x}^{*b} wird B mal durchgeführt, wobei die Zahl B der Bootstrap-Stichproben im Falle der Schätzung des Standardfehlers üblicherweise ca. 50-200 beträgt; bei anderen Anwendungen ist ein höherer Rechenaufwand notwendig, z. B. muss bei Bootstrap-Konfidenzintervallen mit einem 10-fach höheren Aufwand gerechnet werden (Efron, 1993). Die Stichprobenfunktion $t(\mathbf{x})$ wird bei jeder Bootstrap-Stichprobe ermittelt. Die Berechnung dieser Statistik t für jedes \mathbf{x}^{*b} wird als *Bootstrap-Wiederholung* bezeichnet. Die Bootstrap Schätzung der Standardfehlers der Statistik $t(x_1, \dots, x_n)$ berechnet sich als Standardabweichung der Bootstrap Wiederholungen (Efron, 1993 S. 13 Gl. 2.3)

$$\hat{se}^{boot} = \left(\frac{1}{B-1} \sum_{b=1}^B [t(\mathbf{x}^{*b}) - t(\bullet)]^2 \right)^{1/2} \text{ mit } t(\bullet) = \frac{1}{B} \sum_{b=1}^B t(\mathbf{x}^{*b}) \quad (\text{A2.2})$$

Abb. A2.1 zeigt das Bootstrap-Prinzip zur Berechnung des Standardfehlers von Stichprobenfunktionen.

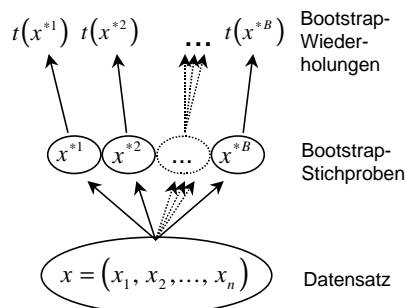


Abb. A2.1: Bootstrap Schema zur Schätzung des Standardfehlers einer Statistik $t(\mathbf{x})$

Vom Originaldatensatz werden B voneinander unabhängige Bootstrap-Stichproben generiert. Jede Bootstrap-Stichprobe hat den Umfang n und wird durch Ziehen mit Zurücklegen gewonnen. Die Bootstrap Wiederholungen werden durch Berechnung der Statistik t von den Bootstrap-Stichproben berechnet. Die Standardabweichung der B Bootstrap-Wiederholungen ist die gesuchte Schätzung des Standardfehlers der Statistik $t(\mathbf{x})$.

Allgemein können die Argumente x_i der Statistik t auch komplizierter sein, z. B. Vektoren, Karten, Bilder etc. (Efron, 1993). Das Bootstrap-Prinzip kann dann an solche Datenstrukturen angepasst werden. In Abb. A2.2 ist das Bootstrap-Prinzip für Ein-Stichproben-Probleme allgemein dargestellt.

Die einfachste Bootstrap-Schätzung des Vorhersagefehlers von Modellen (Efron, 1993)

$$err(\mathbf{X}, F) \equiv \hat{err}^{boot} = E_{0F} \left\{ Q \left[Y_0, \eta_{\mathbf{X}}(U_0) \right] \right\} \quad (\text{A2.3})$$

berechnet sich prinzipiell in analoger Weise wie die des Standardfehlers der Statistik $t(\mathbf{x})$ (Efron, 1993). Zunächst werden, wie oben beschrieben, B Bootstrap Stichproben

$$\mathbf{X}^{*b} = (\mathbf{x}_1^*, \mathbf{x}_2^*, \dots, \mathbf{x}_n^*), \text{ mit } b = 1, \dots, B \quad (\text{A2.4})$$

vom Umfang N durch n -faches, zufälliges Ziehen mit Zurücklegen vom Originaldatensatz (Efron,

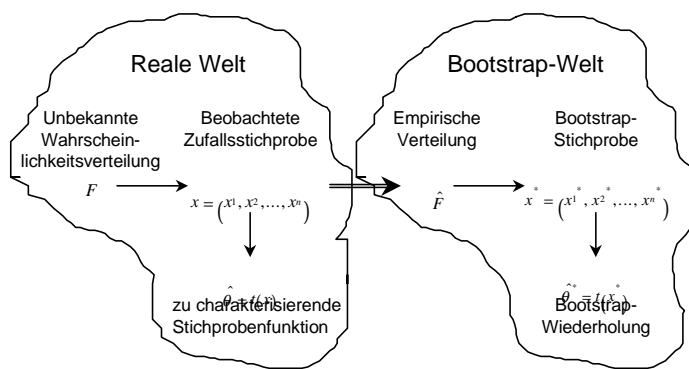
$$1993) \mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_n \\ y_1 & y_2 & \dots & y_n \end{pmatrix}, \text{ mit } \mathbf{u}_i = \begin{pmatrix} u_{li} \\ \vdots \\ u_{pi} \end{pmatrix} \quad (\text{A2.5})$$

gebildet und das Modell $\eta(\mathbf{u})$ an jede dieser Bootstrap-Stichproben \mathbf{X}^{*b} angepasst und optimiert:

$$\eta_{\mathbf{X}^{*b}}(\mathbf{U}^{*b}) = \hat{\mathbf{y}}^{*b} \text{ mit } \sum_{i=1}^n Q(y_i^{*b}, \eta_{\mathbf{X}^{*b}}(\mathbf{u}_i^{*b})) = \min \quad (\text{A2.6})$$

Abb. A2.2: Allgemeines Bootstrap-Prinzip für Ein-Stichproben-Probleme.

Das Bootstrap-Prinzip besteht im Wesentlichen darin, analog wie für die Original-Stichprobenentnahme in der 'realen Welt' eine 'Bootstrap-Welt' zu bilden. In der 'realen Welt' wurde eine Stichprobe \mathbf{x} durch zufällige Beprobung aus einer unbekannten Verteilung F erzeugt. Auf Basis der Stichprobe kann jedoch nur eine einzelne Stichprobenfunktion $\hat{\theta} = t(\mathbf{x})$ berechnet werden, deren statistische Eigenschaften aber gesucht sind,



z. B. der Standardfehler $se_F(\hat{\theta})$. Die 'Bootstrap-Welt' stellt eine Möglichkeit dar, beliebig viele Wiederholungen der Bootstrap-Statistik $\hat{\theta}^*$ zu erzeugen, so dass statistische Aussagen über $\hat{\theta}$ gemacht werden können. Der entscheidende Schritt des Bootstrap-Prinzips ist daher die Berechnung einer Verteilung \hat{F} auf Basis der Stichprobe aus der unbekannten Funktion F . Jeder andere Teil der Bootstrap-Welt ist analog zur Realen Welt definiert: So wie die Stichprobe \mathbf{x} durch

Zufallsstichprobe aus der Verteilung F entsteht, entsteht \mathbf{x}^* durch zufälliges Ziehen aus \hat{F} . Genauso verhält es sich mit $\hat{\theta}$ und $\hat{\theta}^*$, beide berechnen sich durch Anwendung der Funktion $t(\mathbf{x})$ auf \mathbf{x} bzw. \mathbf{x}^* (Efron, 1993).

Im nächsten Schritt werden diese Modelle auf den Originaldatensatz angewendet um B Schätzungen des Vorhersagefehlers zu erhalten. Die einfachste Bootstrap-Schätzung des Vorhersagefehlers berechnet sich dann als Mittelwert der B Schätzungen des Vorhersagefehlers der Bootstrap-Stichproben:

$$err(\mathbf{X}, F) = E_{0F} \left\{ Q \left[Y_0, \eta_{\mathbf{X}}(U_0) \right] \right\} \equiv \hat{err}^{boot} \equiv \frac{1}{Bn} \sum_{b=1}^B \sum_{i=1}^N Q \left[y_i, \eta_{\mathbf{X}^{*b}}(\mathbf{u}_i) \right] \quad (\text{Efron, 1993}) \quad (\text{A2.7})$$

Im speziellen Fall eines Regressionsmodells ergibt sich (Efron, 1993):

$$\hat{err}^{boot} \equiv \frac{1}{Bn} \sum_{b=1}^B \sum_{i=1}^N \left[y_i - \eta_{\mathbf{X}^{*b}}(\mathbf{u}_i) \right]^2 \quad (\text{A2.8})$$

Das Bootstrap-Prinzip zur Schätzung des Vorhersagefehlers am Beispiel eines Regressionsmodells ist in Abb. A2.3 dargestellt.

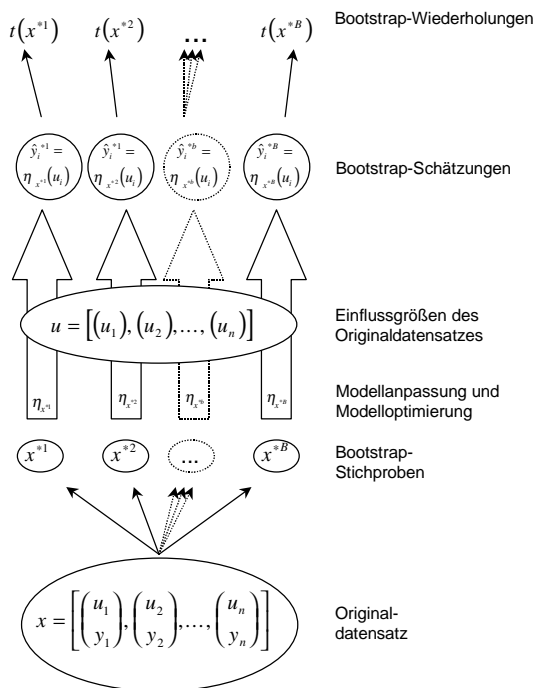


Abb. A2.3: Bootstrap-Prinzip zur Schätzung des Vorhersagefehlers am Beispiel eines Regressionsmodells

Der Algorithmus ist ähnlich wie der zur Schätzung des Standardfehlers von Stichprobenfunktionen. Es werden zunächst auf gleiche Weise voneinander unabhängige Bootstrap-Stichproben vom Originaldatensatz gewonnen. Dann wird an jede Bootstrap-Stichprobe ein Regressionsmodell angepasst, optimiert und auf die Einflussgrößen des Originaldatensatzes angewendet. Dabei ergeben sich B Bootstrap-Schätzungen des Vorhersagefehlers. Der Erwartungswert dieser B Schätzungen ist die gesuchte Bootstrap-Schätzung des Vorhersagefehlers

Diese einfachste Bootstrap Annäherung zur Berechnung des Vorhersagefehlers erweist sich zwar nicht als besonders gut, da die Schätzung des 'scheinbaren Fehlers' $err(\mathbf{X}, \hat{F})$ verzerrt

ist. Sie lässt sich aber mit verschiedenen, nicht

näher ausgeführten, Ansätzen verbessern, so dass man schließlich einen potenten Schätzer für den Vorhersagefehler, den sogenannten 0.632-Bootstrap-Schätzer (Efron, 1993) erhält, auf dessen Herleitung ich hier nicht detailliert eingehe.

Beim 0.632 Bootstrap-Schätzer wird der Bias ebenfalls durch die Differenz zwischen zwischen 'wahrem Fehler' und 'scheinbarem Fehler' geschätzt, hierbei wird jedoch der 'wahre Fehler' auf eine modifizierte Weise ('der wahre Fehler' wird nur durch die Bootstrap-Stichproben geschätzt, bei denen die aktuelle Stichprobe i nicht enthalten ist) geschätzt, außerdem wird diese Biasschätzung um den theoretisch ableitbaren Faktor 0.632 gekürzt. Dieser Faktor $(1 - e^{-1} \approx 0.632)$ ergibt sich als Wahrscheinlichkeit bei $N \rightarrow \infty$, dass bei einem Datensatz von N Beobachtungen eine bestimmte Beobachtung in einer Bootstrap-Stichprobe des Umfangs N enthalten ist. Die allgemeine Formel dieses Schätzers lautet (Efron, 1993):

$$err(\mathbf{X}, F) \equiv \hat{err}^{0.632boot} = 0.368 \frac{1}{N} \sum_{i=1}^N Q[y_i, \eta_{\mathbf{X}}(\mathbf{u}_i)] + 0.632 \frac{1}{N} \sum_{i=1}^N \sum_{b \in C_i} \frac{1}{B_i} Q[y_i, \eta_{\mathbf{X}^{*b}}(\mathbf{u}_i)] \quad (\text{A2.9})$$

Für ein Regressionmodell ergibt sich der Vorhersagefehler zu

$$err(\mathbf{X}, F) \equiv \hat{err}^{0.632boot} = 0.368 \frac{1}{N} \sum_{i=1}^N [y_i - \eta_{\mathbf{X}}(\mathbf{u}_i)]^2 + 0.632 \frac{1}{N} \sum_{i=1}^N \sum_{b \in C_i} \frac{1}{B_i} [y_i - \eta_{\mathbf{X}^{*b}}(\mathbf{u}_i)]^2 \quad (\text{A2.10})$$

wobei C_i jeweils die Menge von Indizes derjenigen Bootstrap-Stichproben darstellt, die den i -ten Datenpunkt nicht enthalten; B_i bezeichnet dementsprechend die Anzahl dieser Bootstrap-Stichproben.

Der Bootstrap ist ein einfacher Algorithmus, der auch bei sehr komplizierten Stichprobenfunktionen, die weit über die klassische Stichprobenstatistik hinausgehen, angewendet werden kann. Es sind dabei keine theoretischen Berechnungen notwendig, das Verfahren ist einfach und vollautomatisch (d. h. nicht interaktiv) als Computerprogramm zu implementieren.

1.3 Jackknife

Das Jackknife-Verfahren wurde erstmals von Quenouille 1956 zur Schätzung des Bias vorgeschlagen, Tukey (1958) erkannte die Möglichkeit damit Standardfehler von Stichprobenfunktionen zu bestimmen und gab dem Verfahren den Namen.

Das Verfahren verfolgt den gleichen Zweck wie das Bootstrap-Verfahren, hat aber Ähnlichkeiten mit der Kreuzvalidierung in dem Sinne, dass reihum eine Beobachtung ausgelassen wird und mit den verbliebenen Werten ein Funktional berechnet wird. Im Gegensatz zur Kreuzvalidierung, bei der der ausgelassene Wert *prognostiziert* werden soll, wird jedoch beim Jackknife ein Parameter θ oder eine Zufallsvariable R der Stichprobe nichtparametrisch geschätzt. Als Beispiele werden bei Efron (1983) die Schätzung des Erwartungswertes, eines Quantils, der Korrelation, des Verhältnisses zweier Zufallsvariablen, Bias, Varianz und des Expected Excess Errors genannt. Zunächst werde ich kurz in die Notationen und Definitionen einführen.

Gegeben sei eine Stichprobe

$$\mathbf{x} = (x_1, x_2, \dots, x_N) \quad (\text{A2.11})$$

und eine Stichprobenfunktion $\hat{\theta} = t(\mathbf{x})$. Gesucht ist der Bias und die Standardfehler von $\hat{\theta}$. Der Jackknife verwendet sogenannte Jackknife-Stichproben, bei denen jeweils eine Beobachtung der Originalstichprobe ausgelassen wird:

$$\mathbf{x}_{(i)} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_N) \text{ mit } i = 1, 2, \dots, N \quad (\text{A2.12})$$

Die Stichprobenfunktion der Jackknife-Stichprobe $\hat{\theta}_{(i)} = t(\mathbf{x}_{(i)})$ heißt analog wie beim Bootstrap Jackknife-Wiederholung von $\hat{\theta}$. Die Jackknife-Schätzung des Bias von $\hat{\theta}$ (Efron, 1983),

$$\hat{bias}^{jack} \equiv E_F \theta(\hat{F}) - \theta(F) \quad (\text{A2.13})$$

ist definiert durch (Efron, 1993):

$$\hat{bias}^{jack} = (N-1)(\hat{\theta}_{(\cdot)} - \hat{\theta}), \text{ mit } \hat{\theta}_{(\cdot)} = \sum_{i=1}^N \hat{\theta}_{(i)} / N \quad (\text{A2.14})$$

Die Jackknife-Schätzung des Standardfehlers einer Stichprobenfunktion ist bestimmt durch (Efron, 1993):

$$\hat{se}^{jack} = \left[\frac{N-1}{N} \sum_{i=1}^N (\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)})^2 \right]^{\frac{1}{2}} \quad (\text{A2.15})$$

Jackknife-Bias korrigierte Stichprobenfunktionen (Efron, 1983)

$$\tilde{\theta} \equiv \hat{\theta} - BIAS = N\hat{\theta} - (N-1)\hat{\theta}_{(\cdot)} \quad (\text{A2.16})$$

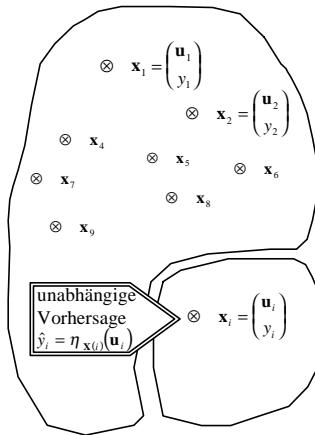
haben im Vergleich mit der Stichprobenfunktion $\hat{\theta}$ einen Bias, der umgekehrt zum Quadrat der Stichprobenanzahl abnimmt, während der Bias der Schätzfunktion $\hat{\theta}$ nur umgekehrt zur Stichprobenanzahl abnimmt (Efron, 1983).

Der Jackknife stellt ein Spezialfall der Einflussfunktion dar (Huber, 1981) und kann anschaulich als einfach zu berechnende Stichprobenkonkretisierung der Einflussfunktion (siehe Kap. 3.1.1.2) aufgefasst werden (Hampel, 1986).

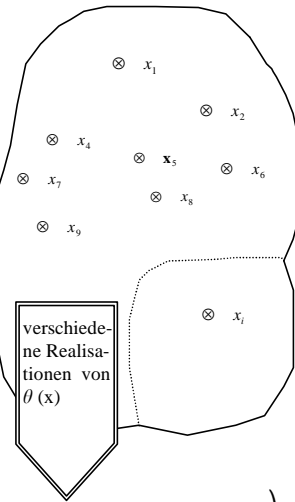
Der Jackknife liefert oft eine einfache, gute und rechentechnisch weniger aufwändige Näherung zur Bootstrap-Schätzung des Bias' und des Standardfehlers, er kann jedoch auch vollständig versagen, wenn die Stichprobenfunktion θ keine differenzierbare Funktion von x ist, d. h. wenn kleine Änderungen im Datensatz große Änderungen der Stichprobenfunktion zu Folge haben können. Ein einfaches Beispiel einer derartigen, nicht glatten, Funktion ist der Median (Efron, 1993).

1.4 Schematischer Verfahrensvergleich

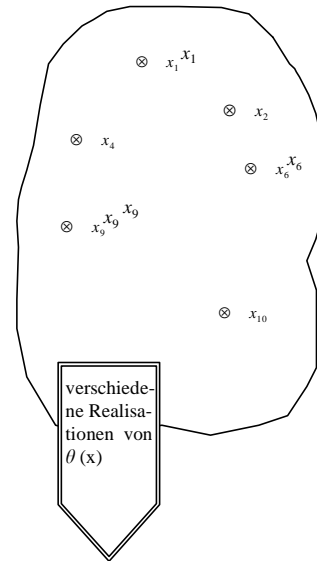
Kreuzvalidierung



Jackknife



Bootstrap



$$\theta_{(i)} = \theta(x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$$

$$\theta_i^* = \theta(x_1, x_1, x_2, x_4, x_6, x_6, x_9, x_9, x_{10})$$

Abb. A2.4: Methodenvergleich der Verfahren Kreuzvalidierung, Jackknife und Bootstrap.

Dargestellt ist jeweils ein Schritt (eine Durchführung), d. h. bei der Kreuzvalidierung die Schätzung eines aus dem Ensemble herausgenommenen Punktes und bei Jackknife und Bootstrap die Realisation einer von n Jackknife- bzw. B Bootstrap Stichproben. Die Kreuzvalidierung hat gegenüber dem Jackknife- und Bootstrap-Verfahren eine ganz andere Zielsetzung. Bootstrap- und Jackknife-Verfahren dienen zur Schätzung von verschiedenen Eigenschaften eines Parameters einer Grundgesamtheit bei unbekannten Verteilungen, d. h. sie ermöglichen es, Statistik mit verschiedenen Realisationen einer Stichprobenfunktion zu betreiben. Die Kreuzvalidierung steht hingegen im Zusammenhang mit einem Modell mit einer Zielgröße und einer oder mehreren Eingangsgrößen. Das Ziel der Kreuzvalidierung besteht in einer Modellüberprüfung, bei der eine Hinzunahme eines neuen Datenpunktes simuliert wird.

Anhang 3: Ergänzungen zur Robusten Statistik

1.1 Problemstellung und Motivation

In der Praxis treten oft Abweichungen von strengen parametrischen Modellvorstellungen über die Verteilungen von Messwerten auf (Hampel et al. 1986, S 21). Die Abweichungen können verschiedene Ursachen haben: Das Auftreten von Fehlern bei der Datengewinnung und -aufbereitung (Messung, Aufschrieb, Runden, Gruppieren, Fehlerfortpflanzung, Berechnung etc.). Derartige Fehler können sowohl ein größeres (sog. Ausreißer) als auch kleineres Ausmaß annehmen und in der Menge guter Daten verborgen bleiben, ein einziger nicht bemerkter Fehler kann jedoch eine statistische Analyse vollständig verderben, insbesondere, wenn ein kleinsten-Quadrate-Algorithmus verwendet wird. Ein Anteil von mehreren Prozent grober Fehler sind in der Wissenschaft verbreitet (Hampel et al., 1986). Moderne robuste Techniken können mit fehlerbehafteten Datensätzen weitaus besser umgehen als objektive oder subjektive Methoden, mit denen versucht wird, bestimmte Datensätze zurückzuweisen.

1.2 Annahmen

Statistische Schlussfolgerungen basieren zum einen auf Beobachtungen, zum anderen auf Annahmen über die den Beobachtungen zu Grunde liegende Sachverhalte. Diese Annahmen erstrecken sich im Wesentlichen auf Zufälligkeit, Unabhängigkeit der Stichprobe und Verteilungsmodelle von unbekannten Parametern. Es wird dabei davon ausgegangen, dass diese Annahmen zwar nicht exakt, aber wenigstens doch annähernd stimmen, es sind mathematisch bequeme Vereinfachungen eines oft unscharfen Wissens oder Glaubens. Diese Vereinfachungen sind unabdingbar notwendig und man geht davon aus, dass kleine Modellabweichungen auch nur kleine Fehler beim Ergebnis bewirken. Leider trifft diese Annahme nicht immer zu. So wurde in den vergangenen Jahrzehnten immer deutlicher, dass die meisten statistischen Prozeduren außerordentlich sensitiv auf kleine Abweichungen vom Modell reagieren, insbesondere wenn Normalverteilung vorausgesetzt wird. Daher wurde die Forderung nach „robusten“ Methoden laut, die weniger sensitiv auf Verstöße gegen die Modellannahmen reagieren.

Auch bei der räumlichen Schätzung ökologischer Parameter werden eine Vielzahl von Annahmen getroffen (hinsichtlich Verteilung, Zufälligkeit und Repräsentativität der Beprobung, etc.), von denen zu erwarten ist, dass sie nicht immer zutreffen. Diese Dissertation will dieser Tatsache Rechnung tragen, deshalb unternehme ich den Versuch, robuste Kriterien zur Beurteilung der Güte von räumlichen Schätzungen zu entwickeln und anzuwenden. In diesem Kapitel gebe ich daher eine kurze Einführung in die wichtigsten Methoden der robusten Statistik. Hierbei beschränke ich mich im Wesentlichen auf das wichtigste Werkzeug der robusten Statistik, der Einflussfunktion mit ihren Derivaten und einigen einfachen robusten Lokationsschätzern. Diese Lokationsschätzer werden im folgenden Kapitel benötigt, um robuste Gütemaße entwickeln zu können. Zunächst will ich den Begriff der Robustheit allgemein definieren und den Unterschied zwischen robusten Methoden und nichtparametrischen bzw. verteilungsfreien Methoden kurz darlegen.

1.3 Definition der Robustheit

Robustheit bezeichnet die Unempfindlichkeit gegenüber kleinen Änderungen der Modellannahmen (Huber, 1981).

Traditionell werden robuste Verfahren im Zusammenhang mit nichtparametrischen und verteilungsfreien Verfahren genannt. Robuste Verfahren können jedoch nicht mit nichtparametrischen oder verteilungsfreien Verfahren gleichgesetzt werden, da robuste Methoden vorwiegend bei parametrischen Verteilungsmodellen angewendet werden und nichtparametrische und verteilungsfreie Verfahren z. T. wenig robust sind (Huber, 1981). Robuste Methoden sind daher parametrischen Ansätzen viel näher als nichtparametrischen und verteilungsfreien, sie unterscheiden sich von nichtrobusten, gewöhnlichen Verfahren insofern, als bei robusten Verfahren die Voraussetzungen parametrischer Verfahren nicht als stillschweigend erfüllt betrachtet werden und dies formal Berücksichtigung findet (Huber, 1981).

Folgende nachfolgende Verfahren und Prozeduren konkurrieren mit robusten Verfahren:

Bei *nichtparametrische* Verfahren werden unabhängig vom Verteilungstyp und Verteilungsfunktion Aussagen über einen unbekannten Parameter gemacht, d. h. sie können bei allen möglichen Verteilungen angewendet werden (Sachs, 1997). Mittelwert und Median einer Stichprobe sind beispielsweise nichtparametrische Schätzer für das Mittel und den Zentralwert der Population. Obwohl nichtparametrisch, ist das Stichprobenmittel sehr ausreißerempfindlich und deshalb nicht robust.

Ein *Test* ist nach der strengen Definition von Huber (1981) *parameterfrei*, wenn die Wahrscheinlichkeit der fälschlichen Ablehnung der Nullhypothese für alle mögliche stetige Verteilungen gleich ist). Ein typisches Beispiel hierfür ist der Zwei-Stichproben-Rang-Test für den Test auf Gleichheit von Verteilungen.

Adaptive Prozeduren wurden erstmals von Stein (1956) beschrieben; sie passen sich selbst an den vorliegenden Datensatz an, so dass z. B. spezifische Lokationsmaße bei hinreichend glatten symmetrischen Wahrscheinlichkeitsdichten asymptotische Effizienz aufweisen (Takeuchi, 1971; Beran, 1974; Sacks, 1975; Stone, 1975). Bei asymmetrischen Verteilungen ist das Verhalten adaptiver Verfahren noch nicht untersucht (Huber, 1981). Ein Übersichtsartikel findet sich bei Hogg (1974) eine Monographie bei Büning (1991). Robuste Methoden unterscheiden sich von adaptiven Methoden dadurch, dass sie eher den Sicherheitsaspekt als den Effizienzaspekt betonen. Adaptive Schätzungen sind nach Stein (1956) und Huber (1981) eher effizienten nichtparametrischen als robusten Verfahren zuzuordnen.

Der Begriff *resistente Prozedur* stammt von Mosteller und Tukey (1977). Sie bezeichnet Insensitivität einer Schätzung oder Teststatistik gegenüber Veränderungen der zu Grunde liegenden *Stichprobe*. Diese Veränderungen können durch kleine Änderungen aller, oder großer Änderungen einiger weniger Stichprobenwerte erfolgen. Obwohl ein deutlich anderer Ansatz vorliegt, sind die Begriffe Resistenz und Robustheit damit für praktische Anwendungen synonym.

1.4 Ausreißerelimination vs robuste Statistik

Ein häufig gegenüber robuster Statistik vorgebrachter Einwand ist, dass robuste Statistik nicht notwendig sei, da ein zweistufiges Datenmanagement mit einer sinnvoll durchgeführten Ausreißerelimination der robusten Statistik ebenbürtig sei. Dabei wird im ersten Schritt der Datensatz durch Anwendung einer Ausreißerregel bereinigt und im zweiten Schritt werden klassische Schätz- und Testverfahren auf die verbliebenen Daten angewendet. Es gibt jedoch drei Gründe, die gegen ein solches Datenmanagement sprechen (Huber, 1981):

1. Bei mehrdimensionalen Regressions-Problemen sind Ausreißer nur sehr schwer zu erkennen. Hampel et al. (1986) zeigen hierzu ein Beispiel.
2. Selbst wenn der vorliegende Datensatz normalverteilt ist und nur einige größere Fehler ‘eingestreut’ sind, sind die bereinigten Daten nicht normalverteilt. Die Situation ist noch viel schlimmer wenn keine Normalverteilung vorliegt, der Datensatz fehlerbehaftet ist, eine Ausreißerelimination durchgeführt wurde und dann eine Normalverteilung angenommen wird. Die Leistungsfähigkeit einer derartigen Vorgehensweise ist weitaus schwieriger zu bestimmen, als die einer robusten Methode. Die klassische Theorie der Normalverteilung ist daher auf *bereinigte* Daten nicht anwendbar
3. Es ist eine empirische Tatsache, dass die besten Ausreißerregeln nicht so leistungsstark sind, wie die besten robusten Prozeduren. Robuste Methoden weisen einen fließenden Übergang zwischen Ablehnung und Verbleib einer Beobachtung im Datenensemble auf und sind deshalb den Ausreißerregeln überlegen. Der fließende Übergang zwischen Ablehnung und Verbleib wird im Unterkapitel ‘Einflusskurven’ näher erläutert.

1.5 Eigenschaften und Ziele robuster Verfahren

Das Verfahren sollte einigermaßen effizient sein. Es sollte so beschaffen sein, dass kleine Abweichungen von den Modellannahmen die Leistungsfähigkeit nur unwesentlich beeinflussen. Größere Modellabweichungen sollten nicht zu beliebigen Werten führen. Die Hauptziele robuster Verfahren sind (Hampel et al, 1986):

1. Datenbeschreibung mit derjenigen Struktur, die am besten mit den Daten übereinstimmt
2. Identifizierung abweichender Datenpunkte (Ausreißer) oder Substrukturen
3. Identifikation von sog. ‘Hebelpunkten’(engl. ‘leverage points’) bei Regressionproblemen, d. h. derjenigen Datenpunkte, die die Regressionsfunktion stark beeinflussen

Im folgenden wird jede Schätzfunktion allgemein mit $T_N = T_N(x_1, \dots, x_N)$ bezeichnet, wenn sie auf einer Stichprobe x_1, \dots, x_N vom Umfang N basiert. Der Grenzwert einer Schätzfunktion T_N für $N \rightarrow \infty$ wird mit T gekennzeichnet. Dabei wird angenommen, dass das Funktional T_N Fisher-konsistent ist (Kallianpur und Rao, 1955), d. h. (Hampel et al., 1986)

$$T(F_\theta) = \theta, \text{ für alle } \theta \quad (\text{A3.1})$$

was bedeutet, dass die Folge der Funktionale $\{T_N; N \geq 1\}$ bei der idealen Modellverteilung asymptotisch die richtige Größe schätzt.

1.6 Kenngrößen der Robustheit: Qualitative und quantitative Robustheit

1.6.1 Qualitative Robustheit

Der Begriff der qualitativen Robustheit wurde von Hampel (1971) eingeführt. Danach ist eine Folge von Schätzern $\{T_N; N \geq 1\}$ *qualitativ robust* auf $F = F_0$, wenn für jedes $\varepsilon > 0$ ein $\delta > 0$ existiert, so dass für alle F und $N \geq N_0$ gilt (Huber, 1981):

$$d_*(F_0, F) \leq \delta \Rightarrow d_*(L_{F_0}(T_N), L_F(T_N)) \leq \varepsilon \quad (\text{A3.2})$$

wobei $L_F(T_N)$ die Verteilung von $\{T_N(F); N \geq 1\}$ und d_* eine Distanzfunktion, z. B. die Prohorov Distanz (Prohorov, 1956), kennzeichnet.

1.6.1 Quantitative Robustheit

Die wichtigsten quantitativen Kenngrößen, welche die Robustheitseigenschaften einer Schätzfunktion T_N beschreiben, sind die Sensitivitätskurve, die Einflussfunktion, gross error sensitivity, local shift sensitivity und der Bruchpunkt.

1.6.2 Sensitivitätskurve

Die Sensitivitätskurve SC einer Schätzfunktion beschreibt die Veränderung einer Schätzfunktion, wenn zu einer Stichprobe x_1, \dots, x_{N-1} noch ein Wert x ergänzt wird (Tukey, 1970):

$$SC(x; x_1, \dots, x_{N-1}; T_N) = N [T_N(x_1, \dots, x_{N-1}, x) - T_{N-1}(x_1, \dots, x_{N-1})] \quad (\text{A3.3})$$

Sie berechnet sich als N -fache Differenz von T_N und T_{N-1} in Abhängigkeit vom Wert x . In Abb. A3.1 sind Sensitivitätskurven für die Schätzfunktionen Median und Mittelwert für einen konkreten Datensatz dargestellt und erläutert. Die Sensitivitätskurve ist zur allgemeinen Beschreibung von Robustheitseigenschaften einer Schätzfunktion nur bedingt geeignet, da sie ja außer von der Schätzfunktion selbst auch von den konkreten Beobachtungswerten und deren Anzahl abhängig ist.

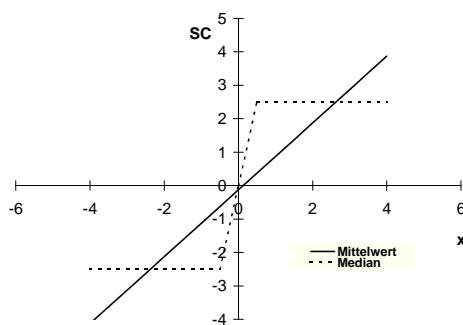


Abb. A3.1: Sensitivitätskurve des arithmetischen Mittels und des Medians

Es liege ein konkreter Datensatz vom Umfang $N = 4$ mit $x_1 = -0.5$, $x_2 = 1.5$, $x_3 = -1$ und $x_4 = 0.5$ vor. Für das arithmetische Mittel ergibt sich mit $T_5 = \bar{x} = \bar{x}_5$ ein

$SC(x; x_1, \dots, x_4; \bar{x}_5) = x - 0,125$. Für den Median ergibt sich

mit $T_5 = \text{med}\{x\} = \text{med}_5$ ein $SC(x; x_1, \dots, x_4; \text{med}_5) =$

$$\begin{cases} -2.5 & x < -0.5 \\ 5x & \text{falls } -0.5 \leq x \leq 0.5 \\ 2.5 & 0.5 < x \end{cases}$$

Die Kurven für den Median und den Mittelwert unterscheiden sich erheblich voneinander. SC ist beim Median beschränkt und für das arithmetische Mittel unbeschränkt. Das arithmetische Mittel kann daher beim vorliegenden Datensatz im Gegensatz zum Median durch Hinzunahme nur eines weiteren Beobachtungswertes x_5 in beliebiger Weise verändert werden. Liegt der neue Wert allerdings zwischen $-0,5$ und $0,5$, reagiert der Median sehr viel heftiger als der Mittelwert, was durch die größere Steigung von SC in diesem Bereich erkennbar ist.

1.6.3 Einflussfunktion

Die Einflussfunktion (IF engl. influence function; gelegentlich wird auch die ältere Bezeichnung Einflusskurve verwendet) ist die verallgemeinerte Form der Sensitivitätskurve, da sie unabhängig von konkreten Beobachtungen und deren Anzahl ist. Sie stellt den Grenzwert der Sensitivitätskurve von Stichprobenwerten aus einer Verteilung mit der Verteilungsfunktion F dar (Hampel, 1968, 1971, 1974):

$$IF(x; F; T) = \lim_{N \rightarrow \infty} SC(x; x_1, \dots, x_{N-1}; T_N) \quad (\text{A3.4})$$

Die Einflussfunktion beschreibt die Reaktion einer Schätzfunktion auf einen zusätzlichen Stichprobenwert, wenn von einer sehr großen Stichprobenzahl aus einer Verteilung F ausgegangen wird. Die Einflussfunktion ist zur allgemeinen Charakterisierung der Robustheit einer Stichprobenfunktion gut geeignet, da sie unabhängig ist von konkreten Beobachtungswerten und deren Anzahl. Die Einflussfunktion stellt ein wichtiges Werkzeug zur Beurteilung von Robustheitseigenschaften dar, von ihr lassen sich weitere wichtige, nachfolgend dargestellte, Größen ableiten, z. B. die *Gross-error sensitivity*, *local-shift sensitivity* und der *rejection point*. Da eine deutsche Übersetzung dieser Begriffe sowohl sprachlich holprig und umständlich als auch weniger prägnant und ungenauer wäre, habe ich bei den nachfolgenden Ausführungen die englischen Fachbegriffe beibehalten.

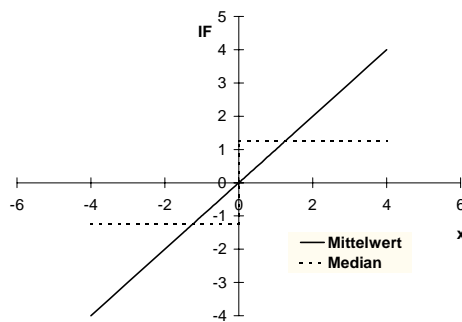


Abb. A3.2: Einflussfunktionen des arithmetischen Mittels und des Medians

Zu Grunde gelegt ist die Standardnormalverteilung mit Mittelwert $\mu = 0$ und Varianz $\sigma^2 = 1$. Bei beliebigen Normalverteilungen ergibt sich für das arithmetische Mittel die $IF(x; F; \bar{x}) = x - \mu$. Beim Median ergibt sich hierfür:

$$IF(x; F; med) = \begin{cases} -\sigma \sqrt{\pi/2} & x < \mu \\ 0 & \text{falls } x = \mu \\ \sigma \sqrt{\pi/2} & x > \mu \end{cases}$$

Die Einflussfunktion des arithmetischen Mittels ist also unbeschränkt, d. h. bei Vorliegen einer normalverteilten Stichprobe kann sich durch die Hinzunahme einer einzigen Beobachtung der Wert der Schätzfunktion beliebig verändern. Die Einflussfunktion des Medians ist hingegen beschränkt, der Wert der Schätzfunktion kann sich im abgebildeten Beispiel nur um $\sqrt{\pi/2}$ verändern. Allerdings weist die Einflussfunktion eine Sprungstelle auf, was bedeutet, dass bereits kleinste Schwankungen des zur Normalverteilung hinzukommenden Wertes eine nichtstetige Veränderung des Wertes der Schätzung zur Folge haben. Arithmetisches Mittel und Median besitzen daher zwar gegensätzliche Vor- und Nachteile, sind aber keineswegs als robust zu bezeichnen.

1.6.4 Gross-Error Sensitivity

Die *Gross-error sensitivity* ist das Supremum des absoluten Betrages der Einflussfunktion für alle x , für die die Einflussfunktion definiert ist (Hampel, 1968, 1974):

$$\gamma^*(T, F) = \sup_x |IF(x; T; F)| \quad (\text{A3.5})$$

Die *gross-error sensitivity* ist ein Maß für den größten Einfluss, den ein einzelner Wert auf eine Schätzfunktion haben kann. Daher ist es erstrebenswert, dass $\gamma^*(T, F)$ finit ist, in diesem Falle heisst T dann *B(bias)-robust* auf F (Rousseeuw, 1981)

1.6.5 Local-Shift Sensitivity

Die *local-shift sensitivity* (Hampel, 1968, 1974) beschreibt die größte Änderung der Schätzfunktion, welche durch eine kleine Änderung einer Beobachtung hervorgerufen werden kann. Kleine ‘Wackler’ von Stichprobenwerten können durch Runden, Gruppieren oder andere kleine Ungenauigkeiten bei der Beobachtung verursacht werden. Diese kleine Verschiebungen von einem Wert x zum Wert y können mathematisch als Hinzunahme einer neuen Beobachtung y und Streichen der Beobachtung x interpretiert werden. Die dabei entstehende maximale Änderung der Schätzfunktion kann daher als Supremum der normierten Differenz der Einflussfunktionen y und x beschrieben werden (Hampel, 1968, 1974):

$$\lambda^* = \sup_{x \neq y} \frac{|IF(y; T; F) - IF(x; T; F)|}{|y - x|} \quad (\text{A3.6})$$

Infinite λ^* können allerdings aufgrund der Normierung auch bei nur begrenzten Änderungen der Schätzfunktion auftreten

1.6.6 Rejection Point

Es wurden bereits sehr früh Versuche unternommen, Ausreißer vollständig zu ignorieren, Schätzer mit dieser Eigenschaft wurden bereits von Daniel Bernoulli um 1769 und Smith 1888 entwickelt (Stigler, 1980). Das Vermögen, Ausreißer vollständig zu vernachlässigen, liegt vor, wenn die Einflussfunktion außerhalb eines finiten Bereiches den Wert Null annimmt. Liegen Beobachtungen mit Messwerten vor, die außerhalb dieses Bereiches liegen, haben diese keinerlei Einfluss auf die Schätzfunktion. Bei Vorliegen einer symmetrischen Verteilung mit Mittelwert Null kann der *rejection point* wie folgt definiert werden (Hampel, 1968, 1974):

$$\rho^* = \inf \{ r > 0 \mid IF(x; T; F) = 0 \text{ und } |x| > r \} \quad (\text{A3.7})$$

Existiert kein derartiges r , ist gemäß der Definition des Infimums $\rho^* = \infty$

1.6.7 Bruchpunkt

Der Bruchpunkt (engl. breakdown point; Hampel, 1968, 1974) ε^* ist in seiner allgemeinen Form ein Maß für die globale Glaubwürdigkeit eines Schätzers, das beschreibt, bis zu welcher Prohorov-Distanz

(Prohorov, 1956) von der angegebenen Modellverteilung eine Schätzfunktion T_N zutreffende Schätzungen liefert (Hampel et al., 1986). Da diese Konzeption in der praktischen Arbeit schwer handhabbar ist, wurden mehrere Bruchpunkt-Versionen entwickelt, die auf endliche Stichproben basieren. Sie sind einfacher, da sie keine Wahrscheinlichkeitsverteilungen enthalten. Im Folgenden verwende ich die Definition von Hampel et al. (1986). Der Bruchpunkt einer Stichprobe ε_N^* einer Schätzfunktion T_N auf Basis der Stichprobe (x_1, \dots, x_N) ist definiert als (Hampel et al., 1986)

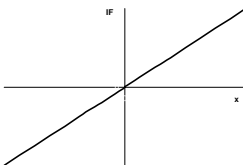
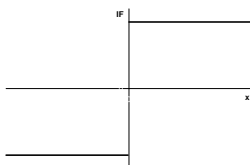
$$\varepsilon_N^*(T_N; x_1, \dots, x_N) = \frac{1}{N} \max \left\{ m \mid \sup_{y_1, \dots, y_m} |T_N(z_1, \dots, z_N)| < \infty \right\} \quad (\text{A3.8})$$

wobei die Stichprobe (z_1, \dots, z_N) aus der Stichprobe (x_1, \dots, x_N) durch Ersetzen von m Beobachtungen x_{i_1}, \dots, x_{i_m} $1 \leq i, j \leq N$ hervorgeht. In vielen Fällen kann durch Grenzwertbildung $\lim_{n \rightarrow \infty} \varepsilon_N^*$ der allgemeine Bruchpunkt ε_N^* gewonnen werden. Der Bruchpunkt einer Stichprobe kann als maximaler Anteil von Ausreißern in einer Stichprobe verstanden werden, ohne dass sich die Schätzfunktion unbeschränkt ändert. Der größtmögliche Wert des Bruchpunktes einer Stichprobe ist 0,5 (Hampel et al., 1986).

1.6.8 Robustheitseigenschaften des arithmetischen Mittels und des Medians

In diesem Kapitel vergleiche ich die beiden gebräuchlichen Lageschätzer arithmetisches Mittel und Median mit Hilfe der oben dargestellten Robustheitseigenschaften; dieser Vergleich soll die Einführung robuster Lageschätzer im nächsten Kapitel motivieren. In Tab. A3.1 sind die wichtigsten Eigenschaften der klassischen Lageschätzer Mittelwert und Median zusammengefasst.

Tab. A3.1: Vergleich der Robustheitseigenschaften des Mittelwertes und des Medians Der Mittelwert ist besonders empfindlich, wenn Beobachtungen mit großen Fehlern behaftet sind, was auf den Median jedoch keine Auswirkungen hat. Genau umgekehrt verhält es sich, wenn in der Stichprobe kleinere Fehler vorkommen: Der Mittelwert reagiert schwach auf kleinere Messwertschwankungen, der Median ist außerhalb des Zentralwertes von solchen Änderungen unbeeinflusst. Im Bereich des Zentralwertes kann jedoch schon eine einzige kleine Messwertschwankung zu einer sprunghaften Änderung der Schätzfunktion führen. Beide Lageparameter sind nicht in der Lage, Messwerte außerhalb eines bestimmten Bereiches (Ausreißer) prinzipiell zu ignorieren. Beim Mittelwert kann sich bereits ein einzelner großer Fehler beliebig stark auf den Schätzwert auswirken, beim Median kann der Anteil an Ausreißern bis zu 50 % betragen, bevor diese sich überhaupt erst auf den zu schätzenden Wert auswirken. Beide Lageschätzer haben zwar einige positive aber auch negative Robustheitseigenschaften und sind deshalb wenig robust.

Eigenschaft	Arithmetisches Mittel	Median
Schematisierte Einflussfunktion		
gross-error sensitivity	unbeschränkt	beschränkt
local-shift sensitivity	beschränkt	unbeschränkt
maximaler Sprung der IF bei Normalverteilung	0	$\sigma\sqrt{2\pi}$
rejection point	∞	∞
Bruchpunkt einer Stichprobe	0	0.5

1.6.9 Maximum-Likelihood-(M-)Schätzer, L-, und R-Schätzer

Der Name M-Schätzer (Huber, 1964) stammt von ‘generalized maximum likelihood’, die Schätzer stellen eine Verallgemeinerung der Maximum-Likelihood-Methode dar. Ein skalenabhängiger M-Schätzer T_N für die Lokation ergibt sich generell aus der Lösung der Minimierungsaufgabe

$$\sum_{i=1}^N \rho(x_i - T_N) = \min \quad (\text{A3.9})$$

oder implizit formuliert:

$$\sum_{i=1}^N \psi(x_i - T_N) = 0, \quad (\text{A3.10})$$

wobei ρ eine beliebige Funktion und $\psi(x - \theta) = \frac{\partial}{\partial \theta} \rho(x - \theta)$ ist (Huber, 1981). Beide Gleichungen sind nicht immer äquivalent, dennoch ist Gl. A3.9 nützlich, um Lösungen für Gl. A3.10 zu finden. Aus diesem Grunde wird ψ oft mit dem M-Schätzer gleichgesetzt, der durch diese Funktion definiert wird.

Bei streng monotonem $\psi(x)$ ergibt sich T_N als eindeutige Lösung der Gl. A3.10. Skalenabhängigkeit bedeutet in diesem Zusammenhang, dass T_N von der Wahl der Messskala für die Stichprobenwerte (x_1, \dots, x_N) abhängig ist. Auf skaleninvariante Versionen von M-Schätzern soll später eingegangen werden.

Median und Mittelwert gehören der Klasse der M-Schätzer an, ihre ρ - bzw. ψ -Funktionen lauten:

$$\begin{aligned} T_n = \text{med}(x) & \text{ mit } \rho(x) = |x| & \psi(x) = \text{sgn}(x) \\ T_n = \bar{x} & \text{ mit } \rho(x) = x^2/2 & \psi(x) = x \end{aligned} \quad (\text{A3.11})$$

$\text{sgn}(x)$ bezeichnet dabei die Vorzeichenfunktion Signum:

$$\text{sgn}(x) = \begin{cases} 1 & x > 0 \\ 0 & \text{falls } x = 0 \\ -1 & x < 0 \end{cases} \quad (\text{A3.12})$$

Eine besondere Eigenschaft der M-Schätzer ist es, dass die Einflusskurve eines skalenabhängigen M-Schätzers proportional zur Funktion $\psi(x)$ ist (Huber, 1981). Die Proportionalitätskonstante $C(F)$ ist dabei nur von der zu Grunde gelegten Verteilung abhängig. Beschreibt $f(x)$ die Dichte dieser Verteilung, $F(x)$ die dazugehörige Verteilungsfunktion, und $E_F[T]$ den Erwartungswert der Schätzfunktion T unter der Verteilung F , so gilt (Hartung, 1986):

$$IC(x, F, T) = C(F) \cdot \psi(x - E_F[T]) \quad (\text{A3.13})$$

Die Proportionalität der Einflussfunktion hat zwei Konsequenzen: Zum einen ist bei den M-Schätzern die Form der Einflussfunktion stets unabhängig von der zu Grunde gelegten Verteilung; daher sie sind robust gegenüber Abweichungen von Verteilungsannahmen. Zum anderen können M-Schätzer mit

maßgeschneiderten Eigenschaften leicht entwickelt werden, in dem eine Funktion $\psi(x)$ so gewählt wird, dass sie die gleiche Form besitzt wie eine vorzugebende Einflusskurve mit den gewünschten Eigenschaften (Hartung, 1986).

Falls $\psi(x)$ monoton und beschränkt (aber nicht notwendigerweise stetig) ist, berechnet sich der Bruchpunkt von skalenabhängigen M-Schätzern zu (Huber, 1981)

$$\varepsilon_M^* = \frac{\eta}{1+\eta} \quad \text{mit} \quad \eta = \min \left(-\frac{\psi(-\infty)}{\psi(+\infty)}, -\frac{\psi(+\infty)}{\psi(-\infty)} \right) \quad (\text{A3.14})$$

Wenn $-\psi(-\infty) = \psi(+\infty)$, ergibt sich $\eta = 1$ und damit $\varepsilon^* = 0,5$, d. h. der maximal mögliche Bruchpunkt überhaupt. Bei nicht beschränktem $\psi(x)$ ist der Bruchpunkt von M-Schätzern stets $\varepsilon^* = 0$.

1.6.9.1 Skaleninvarianz

Bislang wurden nur skalenabhängige Versionen von M-Schätzern beschrieben. Normalerweise sind M-Schätzer skalenabhängig. Eine Ausnahme bilden der Median und der Mittelwert, die in der Menge aller M-Schätzer enthalten sind und skaleninvariant sind. Skalenabhängige M-Schätzer können skaleninvariant gemacht werden, in dem bei der Minimierungsaufgabe eine Skalenschätzung S_n mit einbezogen wird. Zur Skalenschätzung bietet sich der Median der Medianabweichung, MAD, an (Hampel et al., 1986), der mit Hilfe des Faktors c so standardisiert wird, dass bei der idealen Modellverteilung Konsistenz vorliegt (Huber, 1981;:

$$S_N = c \cdot \text{MAD}_N = c \cdot \text{med} \{x_i - \text{med} \{x_i\}\}, \quad (\text{A3.15})$$

$c = 1.483$ bei Normalverteilung (Hampel et al., 1986)

Mit dieser Skalenschätzung kann die Lageschätzung durch Lösen der modifizierten Minimierungsaufgabe

$$\sum_{i=1}^N \rho \left(\frac{x_i - T_N}{S_N} \right) = \min \quad (\text{A3.16})$$

skaleninvariant gemacht werden. Die skaleninvarianten Versionen der M-Schätzer weisen die gleichen Eigenschaften wie die skalenabhängigen Schätzer auf: Die Einflusskurve ist proportional zur Funktion ψ , der Bruchpunkt eines skaleninvarianten Schätzers ist das Minimum aus dem Bruchpunkt des zu Grunde liegenden skalenabhängigen M-Schätzers ε_M^* mit gleicher ρ bzw. ψ -Funktion und des Bruchpunktes des Skalenschätzers S_N , $\varepsilon^*(S_N)$ (Hartung, 1986):

$$\varepsilon_{MS_N}^* = \min(\varepsilon_M^*, \varepsilon_{S_N}^*) \quad (\text{A3.17})$$

Bei der Verwendung des MAD als Skalenschätzer, mit dem maximal möglichen Bruchpunkt von $\varepsilon^* = 0,5$, bleibt daher der Bruchpunkt der skalenabhängigen M-Schätzung erhalten.

1.6.9.2 Berechnung von M-Schätzern

Von den Sonderfällen Median und Mittelwert abgesehen, erfordert die Berechnung von M-Schätzern immer den Einsatz iterativer Verfahren. Es soll an dieser Stelle das Newton-Raphson-Verfahren zur Lösung der Gleichung

$$\sum_{i=1}^N \psi \left(\frac{x_i - T_N}{S_N} \right) = 0 \quad (\text{A3.18})$$

erwähnt werden. Da obige Gleichung unter Umständen mehrere Lösungen besitzt, ist es wichtig, einen geeigneten Startwert T_N^0 zu wählen, so dass Gl. A3.16 erfüllt ist. Es hat sich gezeigt, dass der Median diesbezüglich ein geeigneter Startwert T_N^0 darstellt. In der Regel liegt dann schon nach einer Iteration ein befriedigendes Ergebnis vor (Hartung, 1986).

Die Schätzung T_N^p im p -ten Iterationschritt ($p = 1, 2, \dots$) berechnet sich mit der Ableitung ψ' von ψ wie folgt (Hartung, 1986; verändert):

$$T_N^p = T_N^{p-1} + S_N \frac{\sum_{i=1}^N \psi \left(\frac{x_i - T_N^{p-1}}{S_N} \right)}{\sum_{i=1}^N \psi' \left(\frac{x_i - T_N^{p-1}}{S_N} \right)} \quad (\text{A3.19})$$

Stabilisieren sich die Schätzwerte T_N^p , d. h. ist für ein p und ein vorzugebendes $\varepsilon > 0$

$$|T_N^p - T_N^{p-1}| < \varepsilon, \quad (\text{A3.20})$$

so wird die Iteration abgebrochen und $T_N = T_N^p$ ist der gesuchte M-Schätzer.

1.6.9.3 Huber-k-Schätzer

Die Huber-k-Schätzer stellen eine Klasse von M-Schätzern dar, bei denen die ρ bzw. ψ -Funktionen des Mittelwert- und Medianschätzer abschnittsweise kombiniert werden. Sie enthalten daher diese Schätzer als Spezialfälle ($|x_i| \leq k \rightarrow \text{Mittelwert}$; $|x_i| > k \rightarrow \text{Median}$), vereinen aber die Vorteile der beiden Schätzer. Huber-k-Schätzer sind durch folgende Definitionen für ρ bzw. ψ festgelegt (Huber, 1964, Formulierung nach Hartung, 1986):

$$\rho(x) = \begin{cases} x^2/2 & \text{falls } |x| \leq k \\ k \cdot |x| - k^2/2 & \text{falls } |x| > k \end{cases}, \text{ bzw. } \psi(x) = \begin{cases} x & \text{falls } |x| \leq k \\ k \cdot \text{sgn}(x) & \text{falls } |x| > k \end{cases} \quad (\text{A3.21})$$

wobei k eine positive, reelle, Konstante darstellt. Aus dieser Definition geht hervor, dass Huber-k-Schätzer sowohl qualitativ als auch B-robust sind und außerdem gilt folgende Gleichung: $-\psi(-\infty) = \psi(+\infty)$, daher haben Huber-k-Schätzer den maximal möglichen Bruchpunkt von 0,5 (Hampel et al., 1986).

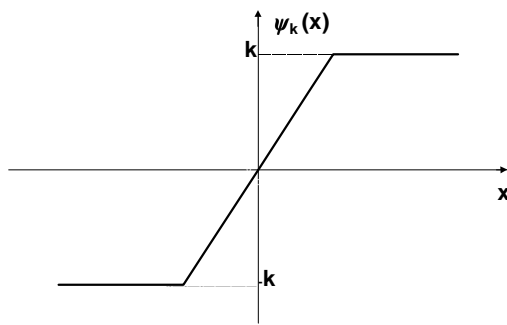


Abb. A3.3 Huber-k-Schätzer definierende ψ – Funktion mit Cut-Off-Punkt k

Dargestellt ist der qualitative Verlauf einer ψ – Funktion, welche einen Huber-k-Schätzer für eine Verteilung mit Erwartungswert 0 definiert. Aufgrund der Eigenschaften der Huber-k-Schätzer ist die den Schätzer definierende ψ – Funktion proportional zur Einflussfunktion des Schätzers. Es wird deutlich, dass Huber-k-Schätzer hinsichtlich der Einflussfunktion eine Synthese zwischen Median und Mittelwert darstellen und die Vorteile beider Verfahren kombinieren: Die Einflussfunktion ist beschränkt weist keine Sprungstellen auf und ist punktsymmetrisch, daher sind sowohl local-shift sensitivity als auch gross-error sensitivity beschränkt; der Schätzer ist somit qualitativ und B-robust mit einem Bruchpunkt von 0,5.

symmetrisch, daher sind sowohl local-shift sensitivity als auch gross-error sensitivity beschränkt; der Schätzer ist somit qualitativ und B-robust mit einem Bruchpunkt von 0,5.

1.6.9.4 Redescending M-Estimators

Bei den Huber-k-Schätzern ist ein beschränkter Einfluss beliebig großer Stichprobenwerte immer vorhanden. D. h. auch klar als Ausreißer erkennbare Beobachtungen haben Einfluss auf die Lageschätzung. Diese Situation wird bei M-Schätzern, deren ψ – Funktion außerhalb eines endlichen Bereiches bis auf Null abnimmt, vermieden. Formal lassen sich diese Schätzer wie folgt beschreiben (Hampel et al., 1986):

$$\psi_r := \left\{ \psi \mid \psi(x) = 0 \text{ für alle } |x| \geq r; 0 < r < \infty \right\} \quad (\text{A3.22})$$

M-Schätzer mit der Eigenschaft, Ausreißer vollständig ignorieren zu können, heißen *redescending M-estimators*. Gelegentlich werden aber von einigen Autoren auch Schätzer redescending genannt, deren ψ – Funktion nur asymptotisch gegen Null abnehmen (z. B. Holland & Welsch, 1977). Es wurden verschiedene ψ – Funktionen für redescending M-estimators vorgeschlagen, so z. B. die linearen *Hampel-Schätzer* (Andrews et al. 1972), die auf der Sinus-Funktion basierenden *Andrew's waves* (Andrews et al. 1972) und die sehr glatte *Biweight-Funktion* von Tukey (Beaton & Tukey, 1974). Eine qualitative Darstellung der entsprechenden ψ – Funktionen wird in Abb. A3.4 gegeben.

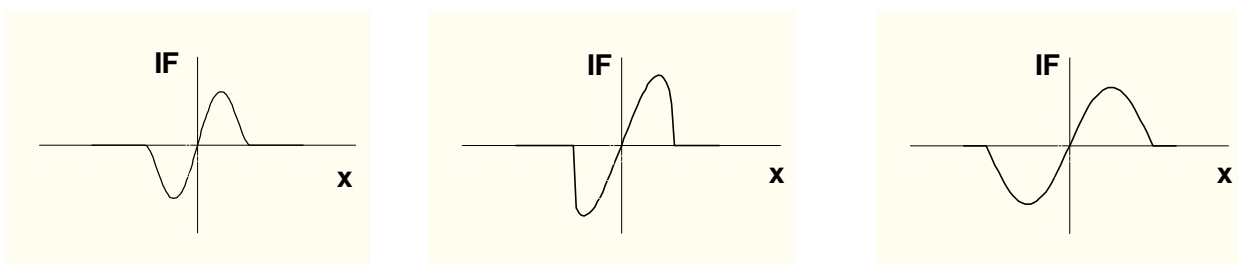


Abb. A3.4: Schematische Darstellung der ψ – Funktionen verschiedener redescending M-estimators

Redescending M-estimators unterscheiden sich von gewöhnlichen M-estimators im Wesentlichen darin, dass die ψ – Funktionen und damit auch die Einflussfunktion außerhalb eines endlichen Bereiches auf Null abfallen. Beobachtungen, die außerhalb dieses endlichen Bereiches liegen, werden auf Grund dieser Eigenschaft vollständig ignoriert und haben keinerlei Einfluss auf die Lageschätzung.

1.6.9.5 L-estimators

Ein L-Schätzer für die Lokation ist eine Linearkombination der geordneten Stichprobe $x_{(1)}, \dots, x_{(N)}$, mit $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(N)}$ (Huber, 1981 vereinfacht):

$$T_N = \sum_{i=1}^N a_{Ni} x_{(i)} \quad (\text{A3.23})$$

wobei die a_{Ni} feste Gewichte darstellen, die von der Stichprobenanzahl abhängen. Je nach Wahl der Gewichte ergeben sich drei wichtige Klassen von L-Schätzern: α -getrimmte Mittel, α -winsorisierte Mittel und α -Gastwirth-Cohen-Mittel (Gastwirth-Cohen, 1970). L-Schätzer sind stets skaleninvariant, ihre Berechnung ist ohne aufwändige Iterationsverfahren möglich. Die Einflusskurve ist jedoch schwerer zu bestimmen.

1.6.9.6 R-estimators

R-Schätzer für die Lokation werden mit Hilfe von Zweistichproben-Rangtests für Lokationsvergleiche gewonnen. Dabei werden zwei Stichproben (x_1, \dots, x_N) und (y_1, \dots, y_N) durch einen Rangtest (z. B. Wilcoxon Rangsummentest) verglichen. Die Stichprobe (y_1, \dots, y_N) wird dabei aus der Originalstichprobe (x_1, \dots, x_N) durch Multiplikation mit -1 und anschließender Addition von $2T_N$ gewonnen:

$$y_i = 2T_N - x_i, \quad \text{für } i = 1, \dots, N$$

Der Wert T_N wird danach so lange optimiert, bis mit Hilfe des Rangtests ein möglichst kleiner Unterschied zwischen (x_1, \dots, x_N) und (y_1, \dots, y_N) festgestellt werden kann. Der auf diese Art gewonnene Wert T_N ist der R-Schätzer für die Lokation von (x_1, \dots, x_N) .

1.6.9.7 Weiter führende Literatur

Auf weitere Schätzer zur Bestimmung der Lokation, wie z. B. A-, D-, P-, S- und W-Schätzer soll hier nicht eingegangen werden. Hierzu verweise ich auf Hampel et al. (Übersicht; 1986), Lax (A-Schätzer; 1975), Wolfowitz (D-Schätzer; 1957), Johns (P-Schätzer; 1979), Rousseeuw & Yohai (S-Schätzer; 1984) und Tukey (W-Schätzer; 1970).

Anhang 4:

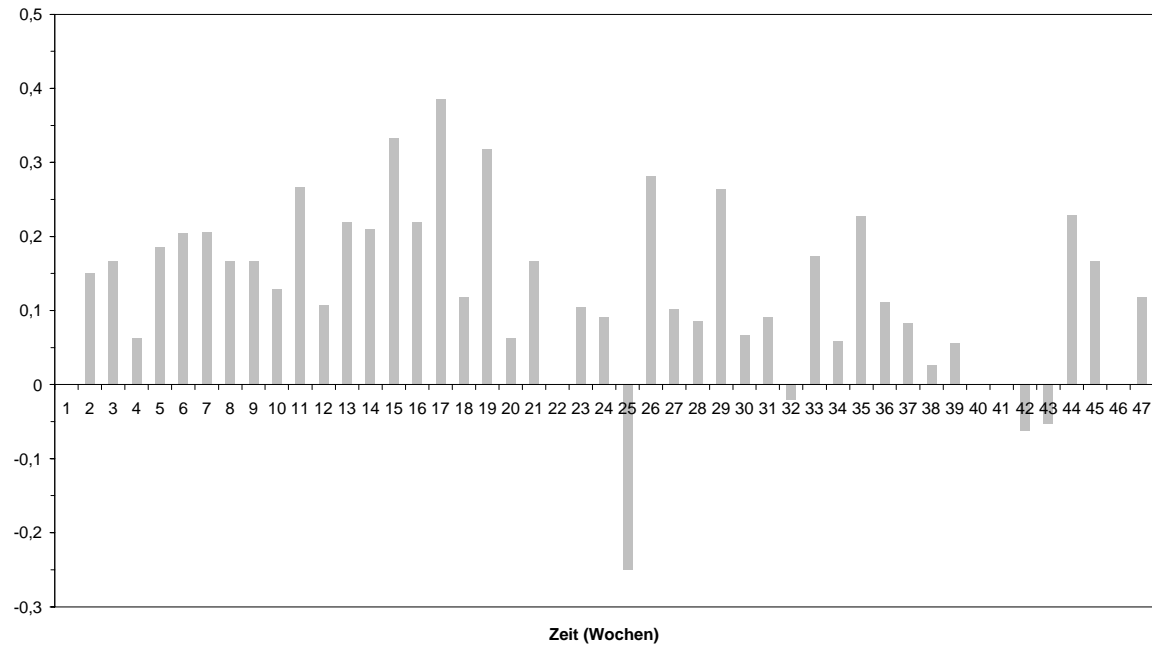
Numerische Kodierung der Bodenschätzungsbodenarten

Die Paare BS-Bodenart – Numerische Kodierung sind zweifach angegeben. Im linken Block alphabetisch nach BS-Bodenart sortiert, im rechten Block in numerisch aufsteigender Reihenfolge

Bodenart nach BS	Numerische Kodierung	Numerische Kodierung	Bodenart nach BS
f3S13	2	1	S13-S14
f3S13-f3S14	3	2	f3S13
f3S14	6	3	f3S13-f3S14
f3S14-(Lfs4)	7	4	S14
Lfs1	34	5	Lfs4-f3S13
Lfs2	25	6	f3S14
Lfs2-(Lfs3)	26	7	f3S14-(Lfs4)
Lfs2-(Lkr3)	27	8	Lfs4-f3S14
Lfs2-Lfs3	32	9	Ls4
Lfs2-Lfs4	21	10	Lfs4
Lfs2-Lkr3	33	11	Lfs4-(f3S14)
Lfs2-Ls3	22	12	Lfs4-(Lfs3)
Lfs3	18	13	Ls3
Lfs3-(Lfs2)	20	14	Lfs3-Lfs4
Lfs3-(Lfs4)	19	15	Ls2-Ls3
Lfs3-f3S14	16	16	Lfs3-f3S14
Lfs3-Lfs4	14	17	Lfs4-Lschlu3
Lfs3-Lkr3	24	18	Lfs3
Lfs3-Lschlu3	28	19	Lfs3-(Lfs4)
Lfs4	10	20	Lfs3-(Lfs2)
Lfs4-(f3S14)	11	21	Lfs2-Lfs4
Lfs4-(Lfs3)	12	22	Lfs2-Ls3
Lfs4-f3S13	5	23	Ls3-Lkr3
Lfs4-f3S14	8	24	Lfs3-Lkr3
Lfs4-Lschlu3	17	25	Lfs2
Lkr3	35	26	Lfs2-(Lfs3)
Lkr3-(Lfs2)	36	27	Lfs2-(Lkr3)
Lkr3-(Lt2)	37	28	Lfs3-Lschlu3
Lkr3-Ls2	29	29	Lkr3-Ls2
Lkr3-Lschw3	38	30	Lschlu3-Lschli3
Lkr3-Lt3	42	31	Ls3-Lt3
Ls2-Ls3	15	32	Lfs2-Lfs3
Ls3	13	33	Lfs2-Lkr3
Ls3-Lkr3	23	34	Lfs1
Ls3-Lt3	31	35	Lkr3
Ls4	9	36	Lkr3-(Lfs2)
Lschlu3-Lschli3	30	37	Lkr3-(Lt2)
Lschw3	40	38	Lkr3-Lschw3
Lschw3-(Lkr3)	41	39	Lschw3-Lkr3
Lschw3-Lkr3	39	40	Lschw3
Lschw3-Lt3	45	41	Lschw3-(Lkr3)
Lschw3-Lt3-Tl3	44	42	Lkr3-Lt3
Lschw4	43	43	Lschw4
Lt3	46	44	Lschw3-Lt3-Tl3
S13-S14	1	45	Lschw3-Lt3
S14	4	46	Lt3
Tl3-T	47	47	Tl3-T

Anhang 5:

Gütemaße der Schätzung der Bodenfeuchte



Koeffizient der mittleren quadrierten Schätzabweichung (*comsre*) der Schätzung der Bodenfeuchte

Zeit: 14.2.1990 - 13.1.1991; Weiherbachgebiet; Schätzverfahren: Bayes-Markow-Kriging; Berechnet nach Daten von Lehmann (1995)

Anhang 6:

Bodenartentransformationsfunktion

von BKA 4 (4. Auflage) in die BKA 3 (3. Auflage)

BKA 4	BKA 3 - Löss/ Kolluvium	BKA 3 - Keuper	BKA 4	BKA 3 - Löss/ Kolluvium	BKA 3 - Keuper
Ls2	Ls2	Ls2	Ts3	Ts4	Ts3
ls4	ls4	ls4	Tt	tl	T
lt2	lu	lt2	Tu2	ltu	tl
lt3	lt2	lt3	Tu4	lil4	ltu
lts	lts	lts	Tus	lu	ltu
lu	lu	lu	Uls	Uls	Uls
Slu	Slu	Slu	Us	Us	Us
Ss	S	S	Ut2	Ul2	Ul2
Su4	Su4	Su4	Ut3	Ul2	Ul3
tl	lt3	tl	Ut4	Ul3	Ul4
Ts2	Ts3	Ts2	Uu	U	U

Anhang 7:

Nitratstickstoffgehalte im Untersuchungsgebiet

Probenahme: Februar 1992; Quelle: LUFA Augustenberg

Probestellen-Nr.	Rechtswert	Hochwert	Betrieb	Kulturart-Nr.	Kulturart	Zwischenfrucht-Nr.	Zwischenfrucht	Vorfrucht-Nr.	Vorfrucht	Düngungsregime-Nr.	Düngungsregime	Hangposition-Nr.	Hangposition	Nmin 0-3 dm [kg/(ha*3 dm)]	Nmin 3-6 dm [kg/(ha*3 dm)]	Nmin 6-9 dm [kg/(ha*3 dm)]
1	3480534	5446277	1	4	W'weizen	27	Brache	6	S'blumen	2	Rind'gülle	3	H'top	23	35	75
2	3480916	5446148	2	4	W'weizen	27	Brache	9	K'mais	1	Schw'gülle	1	H'fuß	65	192	49
4	3481308	5446128	3	12	Stilllegung	27	Brache	4	W'weizen	3	ohne	3	H'top	29	12	13
5	3481826	5446071	4	4	W'weizen	27	Brache	9	K'mais	2	Rind'gülle	1	H'fuß	26	54	76
6	3480351	5445650	5	3	Senf	27	Brache	4	W'weizen	4	Festmist	2	H'mitte	55	72	58
10	3481359	5445728	3	10	Phacelia	27	Brache	4	W'weizen	3	ohne	1	H'fuß	94	40	38
12	3481923	5445642	3	4	W'weizen	27	Brache	11	Z'rüben	3	ohne	3	H'top	19	20	31
13	3481370	5445567	2	27	Brache	27	Brache	9	K'mais	1	Schw'gülle	1	H'fuß	18	11	15
14	3481429	5445465	2	27	Brache	27	Brache	9	K'mais	1	Schw'gülle	3	H'top	33	36	75
16	3480584	5445217	6	16	Rübsen	27	Brache	4	W'weizen	1	Schw'gülle	3	H'top	37	30	29
17	3480897	5445354	7	3	Senf	27	Brache	18	W'gerste	1	Schw'gülle	3	H'top	85	56	26
18	3481151	5445402	7	3	Senf	27	Brache	18	W'gerste	1	Schw'gülle	2	H'mitte	88	106	79
19	3481438	5445208	3	4	W'weizen	27	Brache	11	Z'rüben	3	ohne	1	H'fuß	31	40	25
22	3480383	5444895	8	4	W'weizen	27	Brache	17	Hafer	4	Festmist	2	H'mitte	10	4	3
23	3480615	5444917	9	4	W'weizen	27	Brache	19	Roggen	4	Festmist	3	H'top	20	16	56
24	3480735	5444734	6	4	W'weizen	27	Brache	6	S'blumen	1	Schw'gülle	1	H'fuß	57	24	23
25	3481037	5444651	4	4	W'weizen	27	Brache	9	K'mais	2	Rind'gülle	1	H'fuß	31	27	53
28	3482128	5444672	10	4	W'weizen	27	Brache	9	K'mais	4	Festmist	1	H'fuß	27	49	77
29	3480378	5444384	9	4	W'weizen	27	Brache	6	S'blumen	4	Festmist	3	H'top	18	30	36
30	3480682	5444494	8	18	W'gerste	7	unbek.	18	W'gerste	4	Festmist	1	H'fuß	18	8	12
31	3480684	5444350	8	3	Senf	27	Brache	17	Hafer	4	Festmist	3	H'top	16	3	3
32	3481307	5444476	4	4	W'weizen	27	Brache	9	K'mais	2	Rind'gülle	3	H'top	30	74	51
33	3481539	5444503	10	3	Senf	27	Brache	4	W'weizen	4	Festmist	1	H'fuß	54	53	42
35	3482058	5444274	3	3	Senf	27	Brache	4	W'weizen	3	ohne	1	H'fuß	37	36	35
36	3482066	5444125	11	12	Stilllegung	27	Brache	4	W'weizen	3	ohne	3	H'top	51	96	56
37	3480375	5443873	6	16	Rübsen	16	Rübsen	9	K'mais	1	Schw'gülle	3	H'top	32	20	35
38	3480935	5443934	12	27	Brache	27	Brache	27	Brache	4	Festmist	1	H'fuß	35	24	38
39	3481223	5443969	4	3	Senf	27	Brache	4	W'weizen	2	Rind'gülle	2	H'mitte	18	7	6
40	3481136	5443679	6	10	Phacelia	27	Brache	15	F'erbsen	1	Schw'gülle	3	H'top	50	25	6
43	3481845	5443804	13	12	Stilllegung	12	Stillleg.	12	Stillleg.	7	Unbek.	3	H'top	21	7	8
44	3480761	5443630	6	12	Stilllegung	27	Brache	15	F'erbsen	1	Schw'gülle	3	H'top	31	23	6
45	3480640	5443387	6	4	W'weizen	7	unbek.	14	Gemüse	1	Schw'gülle	1	H'fuß	29	24	61
46	3480850	5443023	14	18	W'gerste	27	Brache	4	W'weizen	3	ohne	3	H'top	12	8	16
47	3481137	5443281	15	3	Senf	27	Brache	4	W'weizen	4	Festmist	1	H'fuß	30	23	18
48	3481124	5443124	14	3	Senf	27	Brache	4	W'weizen	3	ohne	3	H'top	55	18	29
51	3481931	5443603	14	3	Senf	27	Brache	2	S'gerste	3	ohne	3	H'top	18	7	4
52	3481231	5442894	16	3	Senf	3	Senf	9	K'mais	1	Schw'gülle	1	H'fuß	44	51	25
53	3481168	5442592	16	4	W'weizen	27	Brache	9	K'mais	1	Schw'gülle	1	H'fuß	36	98	199
54	3481365	5442678	16	4	W'weizen	27	Brache	9	K'mais	1	Schw'gülle	1	H'fuß	25	72	126
55	3481915	5442886	17	27	Brache	27	Brache	9	K'mais	2	Rind'gülle	2	H'mitte	18	19	19
56	3482014	5444310	3	12	Stilllegung	12	Stillleg.	4	W'weizen	3	ohne	3	H'top	26	17	13
57	3481769	5444817	11	3	Senf	27	Brache	18	W'gerste	3	ohne	3	H'top	12	24	25
59	3481167	5444693	4	3	Senf	27	Brache	4	W'weizen	2	Rind'gülle	3	H'top	42	44	41
60	3480701	5445936	18	4	W'weizen	27	Brache	9	K'mais	4	Festmist	1	H'fuß	45	59	102
65	3481038	5445686	2	8	Gras	27	Brache	4	W'weizen	1	Schw'gülle	1	H'fuß	12	5	5
1001	3480578	5445804	18	3	Senf	3	Senf	4	W'weizen	4	Festmist	3	H'top	60	45	27
1002	3480584	5445733	18	3	Senf	27	Brache	4	W'weizen	4	Festmist	2	H'mitte	67	59	49
1003	3480595	5445649	18	3	Senf	27	Brache	4	W'weizen	4	Festmist	1	H'fuß	63	46	39

